SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(128th Edition)

AUGUST 2023

BY: CENTRAL LIBRARY

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

PREFACE

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

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

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

This issue contains the information published during **August, 2023**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A 3.3kW Modified LLC Resonant Converter for Grid-Tied EV System Under Wide Voltage Range, ***7.Monish Ansari***, ***6.Gaurav Yadav*** & ***3.Mukhtiar Singh***, Electrical, DTU

2. A Compact Wideband SRR-Based C-Shaped Antenna for 5G New Radio Band n258, ***6.Suresh Angadi***, Thennarasan Sabapathy, ***3.N.S. Raghava***, Muzammil Jusoh, Hamsakutty Vettikalladi, Ali M. Almuhlafi and Mohamed Himdi, ECE, DTU

3. A Cross-Domain Semi-Supervised Zero-Shot Learning Model for the Classification of Hyperspectral Images, ***6.Pallavi Ranjan*** and Gautam Gupta, DTU

4. A Fast & efficient Noise-resistant Recommender System using Semantic Ontology and User Networks, ***8.Ayush Tanwar***, ***3.Vinod Kumar***, ***8.Chitraksh Grover*** and ***8.Arhan Relan***, CSE, DTU

5. A New Cosmological Model: Exploring the Evolution of the Universe and Unveiling Super-Accelerated Expansion, ***7.Himanshu Chaudhary***, Dhruv Arora, Ujjal Debnath, G.Mustafa and S. K. Maurya, Mathematics, DTU

6. A novel cobalt(II) acetate complex bearing lutidine ligand: a promising electrocatalyst for oxygen evolution reaction, Deepika Tanwar, Priya Jain, ***6.Deepali Ahluwalia***, Athul Sudheendranath, Sajesh P. Thomas, Pravin P. Ingole and Umesh Kumar, Applied Chemistry, DTU

7. A Review on Community Detection Approach for Recommender Systems, ***7.Akansha Mittal*** and ***3.Anurag Goel***, CSE, DTU

8. A Review on Recent Encoder-Decoder Architectures for Drivable Region Detection and Segmentation, ***6.Dhruv Sharma***, ***3.S. Indu*** and ***3.N. Jayanthi***, Electronics, DTU

9. A single–Port to Dual-Port Reconfigurable 7T SRAM Bit Cell for High Speed, Power Saving and Low Voltage Application, ***6.Bhawna Rawat*** and ***3.Poornima Mittal***, Electronics, DTU

10. A Statistical Analysis Model of Wind Power Generation Forecasting for the Western Region of India, Sulagna M., Piyush H., ***7.1.Vineet S***. and Pankaj R, Electrical, DTU

11. A SURVEY PAPER ON INTELLIGENT INTRUSION DETECTION SYSTEM BASED ON DEEP LEARNING APPROACH, ***7.Prateek Shrivastava*** and ***3.Rajesh Kumar Yadav***, CSE, DTU

12. Adsorption of anionic dye from aqueous environment using surface-engineered Zn/Cu hydroxy double salt-based material: mechanistic, equilibrium and kinetic studies, ***6.Bhamini Pandey***, ***3.Poonam Singh*** and ***3.Vinod Kumar***, A Chemistry, DTU

13. Advanced Intelligent approach for state of charge estimation of lithium-ion battery, ***6.Deepak Kumar***, ***3.M. Rizwan***, & ***3.Amrish K. Panwar***, Electrical & A Physics, DTU

14. Advancement in steam reforming of methanol to produce hydrogen: A review, ***6.Ravindra Kumar*** and ***3.Anil Kumar***, Mechanical and Environment, DTU

15. Aerodynamic Study of Various Tall Building Under Wind Load, ***8.Payal Devi***, ***8.Abhishek Prakash Paswan***, ***8.Hemant Gautam***, ***8.Rahul Kumar Meena*** and ***3.Ritu Raj***, Civil, DTU

16. An Overview on Integration of Artificial Intelligence Tools to Predict the Nature of Chronic Obstructive Pulmonary Disease, ***6.Sourabh Sharma*** and ***3.Yasha Hasija***, Biotechnology, DTU

17. Apple Leaves Disease Detection Using Multilayer Convolutional Neural Network, ***7.Saradindu Das*** and ***3.Abhilasha Sharma***, Software Engineering, DTU

18. Application of Feedforward Neural Network in Portfolio Optimization and Geometric Brownian Motion in Stock Price Prediction, ***8.Anirudh Kumar Aheer***, ***8.Ashish Kumar Pradhan*** and ***3.Dr. R Srivastava***, Applied Mathematics, DTU

19. Assessment of Spatiotemporal Variations in Water Quality of the Urban River Reach, Yamuna, Delhi, ***6.Nibedita Verma***, ***3.Geeta Singh*** and ***6.Naved Ahsan***, Environmental, DTU

20. Attention-free based dual-encoder mechanism for Aspect-based Multimodal Sentiment Recognition, ***8.Pankaj Gupta***, ***8.Ananya Pandey***, ***8.Ajeet Kumar*** and ***3.Dinesh Kumar Vishwakarma***, Applied Physics and IT, DTU

21. Blockchain-Enabled Healthcare Records Management: A Survey of Implementation Strategies, ***7.Divya Kuntal*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

22. Capacitor Trapped Current Fed ZCS-ZVS Push Pull DC-DC converter for renewable energy conversion, ***7.Dhananjay Yadav*** and ***3.Vishal Verma***, Electrical, DTU

23. Compact BERT-Based Multi-Models for Efficient Fake News Detection, ***8.Vineet Rana***, ***8.Vyom Garg***, ***8.Yatharth Mago*** and ***3.Dr. Aruna Bhat***, CSE, DTU

24. Comparative Analysis Of Machine Learning Models For Fake News Classification, ***8.Archit Gupta***, ***8.Arnav Batla***, ***8.Chaitanya Kumar*** and ***3.Dr. Goonjan Jain***, Applied Mathematics, DTU

25. Comparative Analysis of ResNet and DenseNet for Differential Cryptanalysis of SPECK 32/64 Lightweight Block Cipher, ***8.Ayan Sajwan*** and Girish Mishra, Physics, DTU

26. Comparison of Fuzzy-C-Means Method with KMeans Algorithm and Self Organizing Maps using Image Quantization, ***8.Rishabh Chauhan***, ***3.Rajan Yadav*** and ***8.Rajesh Kumar Gupta***, Applied Mathematics, DTU

27. Constraints on the Parameters of Modified Chaplygin-Jacobi and Modified Chaplygin-Abel Gases in f(T) Gravity Model, ***7.Himanshu Chaudhary***, Ujjal Debnath, Tanusree Roy, Sayani Maity and G.Mustafa, Applied Mathematics, DTU

28. Credit Card Fraud Detection using Neural Embeddings and Radial Basis Network with a novel hybrid fruitfly-fireworks algorithm, ***3.Indu Singh***, ***8.Nilesh Aditya***, ***8.Pranjal Srivastava***, Shivanshu Mittal, Taruvar Mittal and ***8.Noel Vaibhav Surin***, CSE, DTU

29. Cyberbullying Detection using Deep Learning Models in Bengali Language, ***3.Rohit Beniwal***, ***8.Shivam jha***, ***8.Smarth Mehta*** and ***8.Riyan dhiman***, CSE, DTU

30. Deep learning algorithms for person re-identifcation: sate-of-the-art and research challenges, ***8.Ankit Yadav*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

31. Deep Learning-based Binary Classification for Spam Detection in SMS Data: Addressing Imbalanced Data with Sampling Techniques, ***3.Dr. Manoj Sethi***, ***8.Naman Tyagi***, ***8.Parmeet Singh Kalsi*** and ***8.Parupalli Atchuta Rao***, CSE, DTU

32. Design and Control of a 2kW Electric Vehicle Charger, ***7.Sushant Chaturvedi*** and ***3.Alka Singh***, Electrical, DTU

33. Design of advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks, ***6.Gebrekiros Gebreyesus Gebremariam***, ***3.J. Panda*** and ***3.S. Indu***, Electronics, DTU

34. Design and Analysis of 4*4-Bit Vedic Multiplier with 32nm CNFET Technology, ***8.Ritik Gupta***, ***8.Rohit Ghughtyal***, ***8.Sai Prateek Mahapatra*** and ***3.Sonal Singh***, Electronics, DTU

35. Detection and Classification of Diabetic and Hypertensive Retinopathy Using CNN & Autoencoder, ***8.Amardeep Singh Kapoor***, ***8.Akshat Jain*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

36. Development and assessment of carboxymethyl tamarind kernel gum-based pH-responsive hydrogel for release of diclofenac sodium, ***6.Priyanka Meena***, ***3.Poonam Singh*** and ***3.Sudhir G. Warkar***, Applied Chemistry, DTU

37. Development of flood susceptibility map using a GIS-based AHP approach: a novel case study on Idukki district, India, ***6.Zohaib Ahmed Khan*** and ***3.Bharat Jhamnani***, Civil, DTU

38. Elucidating the molecular mechanism of ultrasound-enhanced lipase-catalyzed biodiesel synthesis: a computational study, Karan Kumar1 · Pallavi Patro, Unnati Raut, ***6.Varsha Yadav***, Lepakshi Barbora and Vijayanand S. Moholkar, Biotechnology, DTU

39. Empirical validation of feature selection techniques for cross-project defect prediction, ***3.Ruchika Malhotra*** and ***3.Shweta Meena***, Software, DTU

40. Energy matrices and life cycle conversion analysis of N-identical hybrid double slope solar distiller unit using Al2O3 nanoparticle, Dharamveer Singh, Ashok Kumar Yadav, ***3.Anil Kumar*** and ***3.Samsher***, Mechanical, DTU

41. Energy transfer induced colour tunable photoluminescence performance of thermally stable Sm3+/Eu3+ co-doped Ba3MoTiO8 phosphors for white LED applications, ***7.Pooja Rohilla*** and ***3.A. S. Rao***, Applied Physics, DTU

42. Enhanced solar cell efficiency: copper zinc tin sulfde absorber thickness and defect density analysis, K. C. Devendra, Deb Kumar Shah, ***8.Subhash Kumar***, Nawraj Bhattarai, Dipak Raj Adhikari, Khim B. Khattri, M. Shaheer Akhtar, Ahmad Umar, Ahmed A. Ibrahim, Mohsen A. M. Alhamami, Sotirios Baskoutas and O.-Bong Yang, Electrical, DTU

43. Ensemble Transfer Learning for Distinguishing Cognitively Normal and Mild Cognitive Impairment Patients Using MRI, ***8.Pratham Grover***, Kunal Chaturvedi, Xing Zi, Amit Saxena, Tony Jan and Mukesh Prasad, Biotechnology, DTU

44. Environmental sustainability analysis of biofuels: a critical review of LCA studies, ***6.Kulvendra Patel*** and ***3.S. K. Singh***, Environmental, DTU

45. Evaluating Metaheuristic Algorithms and Neural Networks for PID Tuning, ***8.Abhigyan Adarsh***, ***8.Amal Chawla***, ***8.Anirudh Singh*** and ***3.Bharat Bhushan***, Electrical, DTU

46. Evaluating the impact of wind effects on tall buildings with various plan shapes, ***6.Indrajeet Singh***, ***6.Abhishek Prakash Paswan***, ***6.Chitrakshi***, ***3.Ritu Raj***, ***3.S. Anbu kumar*** and ***8.1.Rahul Kumar Meena***, Civil, DTU

47. Evaluation of Various Beams Using Computational Tools , Bhagya Jayant, ***8.1.Rahul Kumar Meena***, ***8.Aastha Singh***, ***8.Agyeya Mishra***, ***8.Navneet Dalal***, ***8.Rajvir Singh***, ***3.Shailendra Kumar*** and ***3.Ritu Raj***, Civil, DTU

48. Exploring the Efficacy of Deep Learning Models for Multiclass Toxic Comment Classification in Social Media Using Natural Language Processing, ***3.Prashant Giridhar Shambharkar***, ***8.Hastej Singh***, ***8.Harshit Raj Raghav*** and ***8.Hilansh Verma***, CSE, DTU

49. Exploring the Impacts of HDPE Microplastics on Growth and Physiological Behavior of Brassica juncea (Mustard Plant), ***6.Megha Bansal***, ***3.Deenan Santhiya*** and ***3.Jai Gopal Sharma***, Biotechnology and Chemistry, DTU

50. Exploring Tidal Force Effects and Shadow Constraints for Schwarzschild-like Black Hole in Starobinsky-Bel-Robinson Gravity, Dhruv Arora, Niyaz Uddin Molla, ***7.Himanshu Chaudhary***, Ujjal Debnath, Farruh Atamurotov and G .Mustafa, Applied Mathematics, DTU

51. Facial Image Pre-processing for Enhanced Ethnicity Classification, ***8.Keshav Goel***, ***8.Hitesh Dhiman***, ***8.Himanshu Yadav*** and ***3.Shailender Kumar***, CSE, DTU

52. Failure investigations of last stage lowpressure steam turbine blade, ***6.Nitin Sehra*** and ***3.Sushla Rani***, Mechanical, DTU

53. FPGA-Based Acceleration of a Neural Network for Symbol Classification with Hardware-Optimised Multiplier Implementation, ***8.Lakshay Chandna***, ***8.Megha Sharma***, ***8.Kushik Agarwal***, ***3.Anup Kumar Mandpura*** and ***3.Himanshu Singh***, Electrical, DTU

54. Graphene oxide/Mn3O4 Nanocomposites for Advanced Electrochemical Biosensor for Fumonisin B1 Detection, ***6.Sandeep Sarpal***, Avinash Kumar Singh, Hema Bhardwaj, ***3.Nitin Kumar Puri*** and Pratima R. Solanki, Applied Physics, DTU

55. Hindi Hit Songs Prediction Using Machine Learning Algorithms, ***3.Rohit Beniwal***, ***8.Satvik Batra***, ***8.Shivank Gupta*** and ***8.Shashi Shekhar***, CSE, DTU

56. Histone Deacetylase 6 as A Putative Target in Alzheimer's Disease Therapeutics, ***6.Shubham Kumar Shrivastav*** and ***3.Pravir Kumar***, Biotechnology, DTU

57. Huntington's Disease Biomarker Discovery: A Computational Approach, ***6.Jasmeer Singh Kalra*** and ***3.Pravir Kumar***, Biotechnology, DTU

58. Hybrid reversible watermarking algorithm using histogram shifting and pairwise prediction error expansion, ***3.Lavi Tanwar*** and ***3.Jeebananda Panda***, ECE, DTU

59. Hydrostatic Pressure Efect on the Thermodynamic Properties of Quantum Wire Under a Crossed Electromagnetic Field, ***8.Pranay Khosla***, ***8.Sakshi Arora***, ***8.Yash Gupta***, ***8.Priyanka*** and ***3.Rinku Sharma***, Applied Physics, DTU

60. Impact of Wind Effects on High-rise Buildings Having Varying Cross Sections, ***8.Devesh Kasana***, ***8.Dhawal Tayal***, ***8.Dhruv Choudhary***, ***3.S. Anbukumar***, ***3.Ritu Raj*** and ***8.Rahul Kumar Meena***, Civil, DTU

61. Implementation of a Deepfake Detection System using Convolutional Neural Networks and Adversarial Training, ***8.Sahithi Bommareddy***, ***8.Tanya Samyal*** and ***3.Sonika Dahiya***, Software Engineering, DTU

62. Investigating the Impact of Visual Attention Models in Face Forgery Detection, ***8.Ankit Yadav*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

63. Isolated DC-DC Converter with Synchronous Rectification for Electric Vehicle Battery Charging, ***7.Awanish Akash*** and ***3.Mini Sreejeth***, Electrical, DTU

64. Legitimate and spam SMS classifcation employing novel Ensemble feature selection algorithm, ***3.Shailender Kumar*** and ***6.Shweta Gupta***, CSE, DTU

65. Link Prediction in Social Network using Gradient Boosting, ***8.Sanket Barman*** and ***8.Sanyam Srivastava***, CSE, DTU

66. Low Supply QFGMOS CCCII with Low-pass and High-pass Filter Applications, ***8.Ujjval Gupta***, ***8.Subrut Shanker Yagnik***, ***8.Vaibhav Gangian*** and ***3.Anupama***, Electrical, DTU

67. Luminescence Studies on Dy3+ Doped Calcium Aluminum Borosilicate (CABS) Glasses for White Light Emission and Applications in w-LEDs, Ravita, Aman Prasad, ***6.Pooja Rohilla***, ***6.Rajat Bajaj***, ***6.Anu***, Rajesh Punia and ***3.Allam Srinivasa Rao***, Applied Physics, DTU

68. Mapping the Field of Research on Entrepreneurial Success: A Bibliometric Study and Future Research Agenda, Varda Sardana and ***6.Shubham Singhania***, USME, DTU

69. Modeling and simulation of material removal characteristics in magnetorheological shear thickening polishing, Zhen Ma, Yebing Tian, Cheng Qian, ***6.Shadab Ahmad***, Zenghua Fan and Zhiguang Sun, Production, DTU

70. Modified Decoupled Sense Amplifier with Improved Sensing Speed for Low-Voltage Diferential SRAM, ***6.AYUSH***, ***3.POORNIMA MITTAL*** & ***3.RAJESH ROHILLA***, DTU

71. Multilabel News Category Classification using Machine Learning, ***8.Roshan Kumar Shah***, ***8.Sahil Kumar*** and ***8.Shashank***, CSE, DTU

72. Multimodal Sarcasm Detection (MSD) in Videos using Deep Learning Models, ***6.Ananya Pandey*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

73. Natural Ester Oil-Based High Permittivity CaCu3Ti4O12 (CCTO) Nanofluids an Eco-friendly Insulating Medium for Power Transformer, Arun Ram Prasath R T, Sankar Narayan Mahato, Nirmal Kumar Roy, Thomas Paramanandam, Mubeen Akhtar Ansari and ***7.1.Joges Chandra Biswas***, Electrical, DTU

74. Natural Language Processing based Automatic Making of Use Case Diagram, ***3.Prof. Shailender Kumar***, ***8.Aryaman, Aryan*** and ***8.Divyank Yadav***, CSE, DTU

75. Network Trafc Analysis in Software-Defned Networking Using RYU Controller, ***8.Shanu Bhardwaj*** and ***3.Ashish Girdhar***, CSE, DTU

76. Nonlinear Regression Analysis of Rut Profile Data for Optimal Data Storage and Efficient Terrain Condition Analysis, ***6.Manoj Kumar Kalra***, ***3.Ashutosh Trivedi***, and Sanjay Kumar Shukla, Civil, DTU

77. Novel Enzymatic Biosensor Utilizing a MoS2/MoO3 Nanohybrid for the Electrochemical Detection of Xanthine in Fish Meat, Prateek Sharma, ***6.Deeksha Thakur*** and ***3.Devendra Kumar***, Applied Chemistry, DTU

78. Object tracking via a Novel Parametric Decisions based RGB-Thermal Fusion, Satbir Singh, Arun Khosla and ***3.Rajiv Kapoor***, Electronics, DTU

79. Observational constraints on teleparallel effective equation of state, S. H. Shekh, ***7.H. Chaudhary***, A. Bouali and A. Dixit, Applied Mathematics, DTU

80. Offensive Hinglish Text Classification Using Longformer, Bigbird & HingMBert, ***8.Yakshit Bansal***, ***8.Sumit Samdarshi***, ***6.Sushant Kumar*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

81. ON A CLASS OF CERTAIN NON-UNIVALENT FUNCTIONS, ***3.S. SIVAPRASAD KUMAR*** AND ***8.POOJA YADAV***, APPLIED MATHEMATICS, DTU

82. Optimal design of renewable energy based hybrid system considering weather forecasting using machine learning techniques, ***6.Bandana Sharma***, ***3.M. Rizwan*** and P. Anand, Electrical, DTU

83. Optimal Shade Dispersion Strategy for Enhanced PV Power Extraction Under PSCs, ***6.Ranjeet Singh***, ***3.Vinod Kumar Yadav*** and ***3.Madhusudan Singh***, Electrical, DTU

84. Optimal threshold selection for segmentation of Chest X-Ray images using opposition-based swarm-inspired algorithm for diagnosis of pneumonia, ***6.Tejna Khosla*** and ***3.Om Prakash Verma***, CSE and Electronics, DTU

85. Parametric Optimization of FSWed Dissimilar Composite Joints of AA7075 and AA6061 Using RSM, ***3.R. S. Mishra*** and ***6.Sumit Jain***, Mechanical, DTU

86. Partially Omnidirectional and Circularly Polarized MIMO Antenna covering sub-6 GHz Band for 5G Fast Plan, Monika Kaushik, ***7.1.Jasdeep K. Dhanoa*** and Mukesh K. Khandelwal, Electronics, DTU

87. Performance Enhancement of a PMSM Drive using ANFISHCC for Electric Vehicles, ***6.Brijendra Sangar***, ***3.Madhusudan Singh*** & ***3.Mini Sreejeth***, Electrical, DTU

88. Physics Based Numerical Model of a Nanoscale Dielectric Modulated Step Graded Germanium Source Biotube FET Sensor: Modelling and Simulation, Amit Das, ***3.Sonam Rewari***, Binod Kumar Kanaujia, S.S. Deswal and R.S. Gupta, ECE, DTU

89. Potential of Green Energy: Current Status in Indian Subcontinent, ***7.1.Nikita Jaiswal***, Ashish Vishwakarma and Kshama Shukla, Civil, DTU

90. Potential of Rhodosporidium toruloides for Fatty Acids Production Using Lignocellulose Biomass, ***8.Sushant Sunder***, ***7.Anshul Gupta***, Rashmi Kataria and Rohit Ruhal, Biotechnology, DTU

91. Power Quality Enhancement using Adaptive Notch Filter in Grid Distribution System, ***8.Chirag Singh***, ***8.Gautam Kumar***, ***8.Himanshu Kumar*** and ***3.Ankita Arora***, Electrical, DTU

92. Reversible Data Hiding scheme for color images based on skewed histograms and cross-channel correlation, ***8.Priyansh Bhatnagar***, ***8.Prateek Tomar***, ***8.Rishabh Naagar*** and ***3.Rajeev Kumar***, CSE, DTU

93. Role of Oxidative Stress and Genetic Polymorphism of Matrix Metalloproteinase-2 and Tissue Inhibitor of Metalloproteinase-2 in COPD, ***6.Poonam Sangwan***, Huma Quasimi, ***3.Jai Gopal Sharma***, Mohammad Anwar Habib and Mohammad Iqbal Alam, Biotechnology, DTU

94. Sarcasm Detection on News Headline Dataset Using Language Models, ***3.Abhilasha Sharma***, ***8.Abhi Uday Pandey*** and ***8.Apoorv Gupta***, Software Engineering, DTU

95. SDL-Net: A combined CNN & RNN Human Activity Recognition model, ***8.Dhruv Gupta***, ***8.Ananya Komal Singh***, ***8.Naman Gupta*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

96. Semantic Consistency for Assuring Reliability of Large Language Models, ***8.Harsh Raj***, Vipul Gupta, Domenic Rosati and Subhabrata Majumdar, IT, DTU

97. Sign Language to Text Conversion using Hand Gesture Recognition, ***8.Tanya Kemkar***, ***8.Vasudha Rai*** and ***3.Dr. Bindu Verma***, IT, DTU

98. Small-Signal Modeling of GaN-BTG MOSFET for Wireless Applications, ***3.M. M. Tripathi*** and ***6.Ajay Kumar***, Electrical, DTU

99. Solar Power Prediction Based on Recurrent Neural Networks Using LSTM and Dense Layer With ReLU Activation Function, ***8.Deepanshu Gupta*** and ***3.Vanjari Venkata Ramana***, Electrical, DTU

100. SPV-wind-BES-based islanded electrical supply system for remote applications with power quality enhancement, ***6.Sombir Kundu***, ***3.Madhusudan Singh*** and Ashutosh K. Giri, Electrical, DTU

101. Television as a Tool of Memory and Identity in Khaled Hosseini's Novels, ***8.Priya Meena*** and ***3.Rajiv Ranjan Dwivedi***, DTU

102. The efectiveness of research and development tax incentives in India: a quasi-experimental approach, ***3.Aakanksha Kaushik***, USME, DTU

103. Third Order Quadrature Oscillator with Arbitrary Phase Shift Generation and Amplitude Modulation Using MDXCCCTA, Garima Shukla and ***3.1.Sajal K. Paul***, Electronics, DTU

104. Towards Accurate Stress Classification: Combining Advanced Feature Selection and Deep Learning, ***8.Lovish Arya***, ***8.Harshal Chowdhary***, ***8.Ishita Agrawal*** and ***3.Indu Sreedevi***, Electronics, DTU

105. Tuning A PID Controller using Metaheuristic Algorithms, ***3.Mini Sreejeth***, ***8.Rishant Kumar***, ***8.Nishant Tripathi*** and ***8.Prateek Garg***, Electrical, DTU

106. Unbalanced Half-bridge Split Capacitor Power Decoupling with Multi-Order Frequency Control for 800V On-board Battery Charger, Tuyen D. Nguyen, Long H. B. Nguyen, Dong Tran, ***7.1.Ravi Nath Tripathi*** and Hai N. Tran, Electrical, DTU

107. Using Dielectric Rods to Reconfigure Combline Filters in Center Frequency and Bandwidth, ***7.1.ABHISHEK SHARMA***, SANTIAGO COGOLLOS, VICENTE E. BORIA, AND MARCO GUGLIELMI, Electronics, DTU

108. VDTA-C based voltage mode Tow-Thomas Biquad filter, ***7.Puneet Pandey***, ***3.Dr. Bhavnesh Jaint*** and ***6.Dr.Garima Mann***, Electrical, DTU

109. Water surface profle prediction in non-prismatic compound channel using support vector machine (SVM), ***7.Vijay Kaushik*** and ***3.Munendra Kumar***, Civil, DTU

1. *Vice Chancellor*
2. *Pro Vice Chancellor*
3. *Faculty*
4. *Teaching-cum-Research Fellow*
5. *Asst. Librarian*
6. *Research Scholar*
7. *PG Scholar*
8. *Undergraduate Student*

1.1. *Ex Vice chancellor*
2.1. *Ex Pro Vice Chancellor*
3.1. *Ex Faculty*
4.1. *Alumni*
5.1 *Others*
6.1. *Ex Research Scholar*
7.1. *Ex PG Scholar*
8.1. *Ex Undergraduate Student*

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

A 3.3kW Modified LLC Resonant Converter for Grid-Tied EV System Under Wide Voltage Range

1st Monish Ansari
Electrical Engineering Department
Delhi Technological University
Delhi, India
monishansari7007@gmail.com

2nd Gaurav Yadav
Electrical Engineering Department
Delhi Technological University
Delhi, India
gauravyadav_2k21phdee17@dtu.ac.in

3rd Mukhtiar Singh
Electrical Engineering Department
Delhi Technological University
Delhi, India
smukhtiar_79@yahoo.co.in

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205690

***Abstract*—This study describes the design of a single-phase, two-stage, On-Board charger (OBC) for light loads that may function in the Grid to Vehicle (G2V) operation mode. The first stage AC-DC converter of EV (Electric Vehicle) charger plays a very important role in supporting the grid power factor unity and the H-bridge inverter is regulated for this purpose using a Second Order Generalised Integrator (SOGI) control approach, which exhibits a stable steady state as well as excellent dynamic behaviour. To maintain the international electrotechnical Commission (IEC)-61000-3-2 standard for the power quality PQ of the grid throughout operations. Additionally, the design's performance has been verified in the presence of Sag and Swell grid voltage and diverse operating conditions. The second stage is the Modified LLC converter is used for DC-DC conversion, this converter is capable of achieving Zero Voltage Switching (ZVS) for all switches. The modified LLC converter uses fewer switches than a conventional LLC converter, which greatly reduces reverse power and turn-off losses while also allowing for significantly higher Voltage conversion. A Proportional Integral controller is implemented to control the DC-DC converter's switches. Using a PWM controller with a steady battery current, the charging current and voltage of the battery are tracked and controlled. For the 3.3kW system ratings, an extensive simulation analysis is performed. To examine the grid's reactivity, an inductive load additionally has been attached to the grid.**

Keywords—EV (Electric Vehicle), Second Order Generalized Integrator, Modified LLC Converter, Active Power Factor Correction (APFC), Phase Locked Loop (PLL).

I. INTRODUCTION

An amazing technology that has always moulded and influenced human culture and propelled social advancement for ages is the electric car [1]. The automobile has been an immensely significant vehicle throughout a large portion of the world [2]. Electric transportation the requirement for practically every form of public conveyance and around 50% of all private for-hire transit to have electric cars by 2030 makes vehicular technology an increasingly prominent field [3]. Therefore, increasing the viability and demand for these e-cars, particularly plug-in hybrid e-cars (PHEVs) and hybrid electric cars (HEVs). The construction of onboard charging infrastructures for EVs is necessary to reduce reliance on such cars. To extend the amount of time and miles that a vehicle may be driven off of a battery pack, plug-in electric vehicles (PEVs) need to overcome the problem of distance anxiety.

The (APFC) stage, which can be powered by mono-phase, ternary-phase, or each source of power, is often built into EV chargers. This stage is followed by several stacks of isolated DC-DC converters that operate via identical input-output dc connections [4]. For several reasons including its larger density of power and a threefold faster charging than a single-phase system, a three-phase system is typically favored. Numerous AC-DC power converter configurations have been suggested in the literature [5] for the APFC stage development. The three-phase voltage source converter (VSC) has the most potential for the APFC stage because of its higher power density, higher efficiency, ability to manage bidirectional power flow, and ease of control adoption.

Moving on to the FEC controller, its major goal is to keep the grid current within the boundaries of the IEC *61000-3-2* standard while keeping an uninterrupted DC-link voltage. In G2V mode, it is also primarily to maintain the source's unity power factor and the grid current's THD below 5% following IEEE standards. The performance of EV chargers is negatively impacted by several grid uncertainties (sag, swell, and harmonics) in the supply. Added to that, the PLL's preciseness in estimating the synchronizing phase degrades under non-ideal (distorted and abnormal) grid voltage circumstances. Under both ideal and non-ideal grid voltage circumstances, SRF and PQ theory are equally beneficial, as suggested in [6]. However, for the PLL to function well under problematic grid voltage circumstances. A substantial advancement in estimating synchronization reference phases is required. A positive sequence extractor-based EPLL [8] and an SSI-PLL [7] are proposed in the published work in this area. The three-phase system's three-phase multiplication procedure and complicated trigonometric calculations are introduced by using EPLL. In [9], a novel strategy known as the SOGI algorithm for synchronization reference phase estimation is put forward. Although the SOGI's transient performance at different frequencies causes significant phase shifts. As a result, [9] proposes a SOGI-FLL based on frequency locking rather than traditional phase locking. FLL does, however, make the controller more complicated and diminish the steady-state margin when added to SOGI.

For high DC-DC power conversion the LLC resonating and Dual Active Bridge (DAB) converters are most suitable choices. Due to the fact that both converters offer inherent soft-switching capabilities and deliver effective energy conversion [10], [11]. In order to increase overall system effectiveness and dependability, Paper [12] describes a bidirectional EV charger using the SOGI technique. PLL could be employed in the control loop to enhance system performance. In [13], a two-way OBC technology has been presented in which G2V, V2G, and V2L operations are studied. To preserve the integrity of the electricity in grids

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:03 UTC from IEEE Xplore. Restrictions apply.

while using the G2V mode of functioning the charging of the EV's batteries needs to be managed.

The LLC converter is a good topological arrangement for the isolated dc-dc conversion stage, as seen in Fig. (1), due to its soft-switching properties for every one of the power components from 0 to full load spectrum [14], [15]. A synchronous rectifier (SR) is typically utilized to identify the current zero-crossing point or sense the device diminished voltage to accomplish a high efficiency of conversion. In unidirectional power applications, LLC converters often utilize these SR approaches. fortunately, when it is employed in bidirectional power applications, the SR has to switch among the two-side bridges following the power direction because of the asymmetry of the topology and gain characteristics. The approach to management will become more complicated as a result of sensing both the power direction and the shifting control logic.

An On-Board PEV charger is suggested in the present investigation with additional characteristics to increase power quality. A shunt compensator is used for crucial current removal using the SOGI approach input to the H-Bridge Inverter. At the DC-DC converter side, the high conversion gain is achieved with a lesser number of switches in comparison to Dual Active Bridge (DAB) converter, additionally, all switches achieved ZVS under the proposed modulation technique. The battery is being charged for different referenced current, and safety and reliability is maintained with the help of the proposed isolated converter. The suggested system features a G2V approach to functioning harmonic elimination and supply-side power factor enhancement. Power batteries made of lithium-ion with a modified LLC DC-DC converter. In addition, the charger works better in ideal and non-ideal conditions. Furthermore, the modified LLC-grade interactive charger also performs well in voltage sag and swell conditions. The hysteresis current controller has the advantage of less complexity and better performance than the other current controller. During the distorted grid voltage condition, the SOGI-based voltage filter out the distortion from the voltage, therefore, is not similar to the measured grade voltage during the distorted grid voltage condition. In the aforementioned piece, a single-phase OBC EV charger concept for extreme voltage and enormous energy charging applications is presented in its entirety.

II. SYSTEM CONFIGURATION

The architecture utilized for EV chargers is depicted in Fig. (1). The system consists of two conversion stages, the initial of which is an AC-DC conversion stage and the other of which is a DC-DC conversion stage. With the suggested controlled arrangement, the first stage AC-DC converter functions as a regulated rectifier. The second stage is a modified LLC converter with six MOSFET switches with ZVS characteristics. The Lithium Ion battery is charged after the second stage with Constant Current (CC) charging.

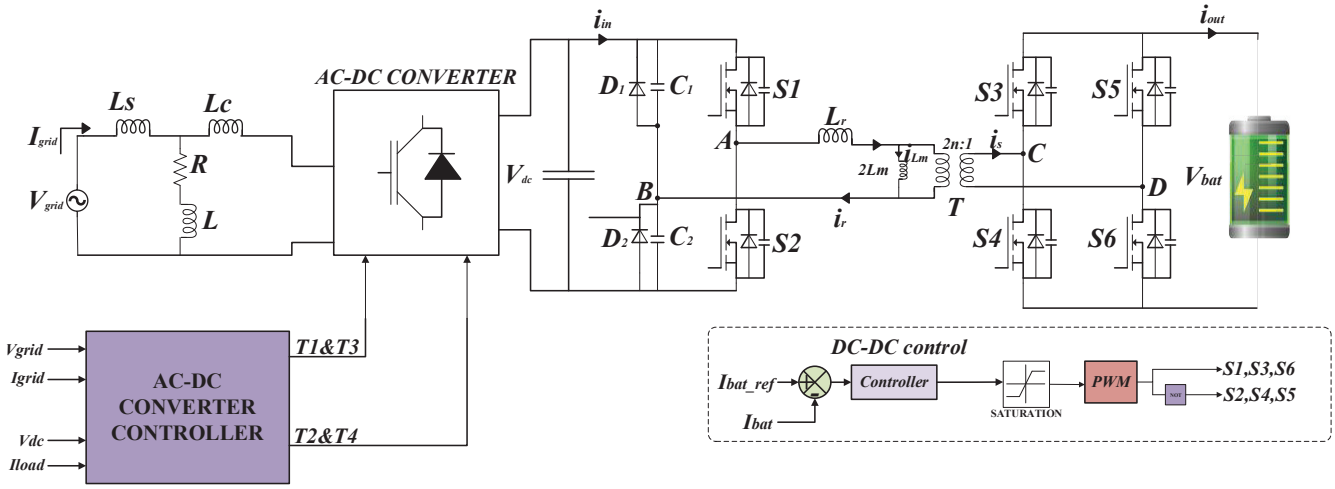


Fig. 1. Circuit Diagram of System configuration for On-Board EV charger (G2V operation).

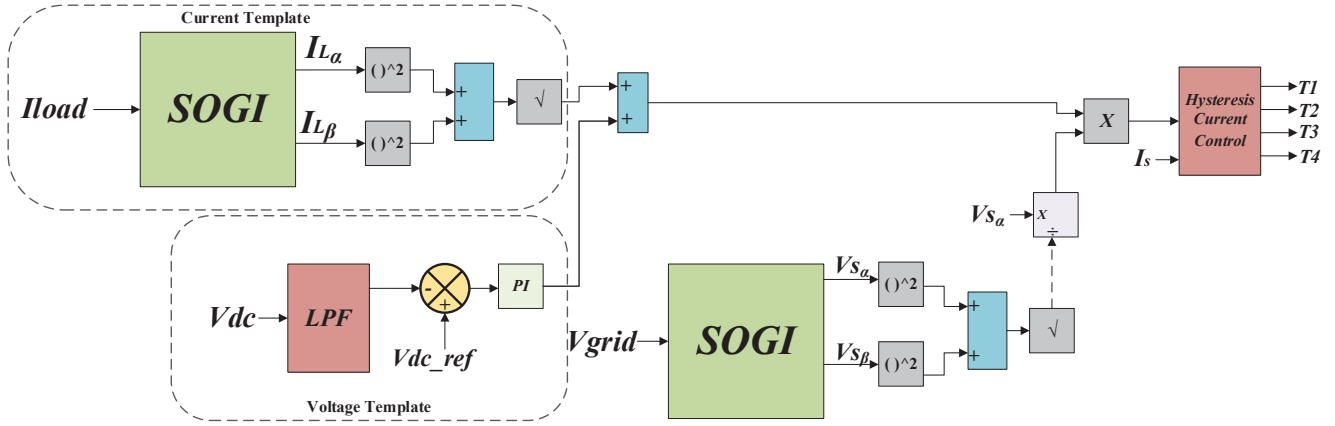


Fig. 2. Control scheme for AC-DC converter stage.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:03 UTC from IEEE Xplore. Restrictions apply.

III. CONTROLLER DESIGN

A. Second-Order Generalized Integrator Method: [SOGI]

The SOGI block's goal, as previously mentioned, is to provide two in-phase, in-quadratic signals to the mains voltage. The percentage of errors is k times increased whenever the supply is first compared to the in-phase signal. This parameter has a substantial impact on the overall efficacy of SOGI. This block adapts the system to shifting frequencies by using the ω` created by the 2nd half of the configuration.

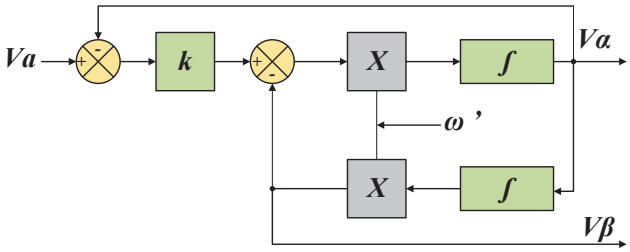


Fig. 3. Representation of SOGI controller.

The aforementioned block can eliminate the 100Hz and 120Hz frequency components from the detected sample. The frequency spectrum arrangement of SOGI is as outlined below:

$$SOGI = \frac{\omega s}{\omega^2 + s^2} \tag{1}$$

ω is the resonance frequency, commonly referred to as the basic angular frequency. For voltages α and β, the closed-loop conversion function is presented in (2) and (3), below:

$$H_\alpha(s) = \frac{v_\alpha(s)}{v(s)} = \frac{k\omega s}{s^2 + k\omega s + \omega^2} \tag{2}$$

$$H_\beta(s) = \frac{v_\beta(s)}{v(s)} = \frac{k\omega^2}{s^2 + k\omega s + \omega^2} \tag{3}$$

$H_\alpha(s)$ is a 2nd-order BPF with no phase shift, while $H_\beta(s)$ is a 2nd-order low-frequency filter with gain equal to 1 and 90° phase departure at 314 rad/s.

By using a SOGI filter, the basic aspect of the load current is removed. The SOGI block processes the supplied signal (I_{load}), which results in the equal-phase and quadrature-phase signals I_α and I_β respectively, being produced at the output. Thus, it is possible to determine the basic current (I_f) using

$$I_f = \sqrt{I_f{}^2 + I_\beta{}^2} \tag{4}$$

In order to determine the switching loss in the shunt compensator using the V_{dc_ref}*, the V_{dc} is traversed through an LPF. Using the inaccuracy of the dc link voltage as input, a PI controller computes the switching loss (I_{loss}) of the recommended system. A steady 400V DC connection voltage is sustained. The formula is described below:

$$I_{loss}(k) = [(k_p + k_i / s)(V_{dc_ref}{}^*(k) - V_{dc}(k))] \tag{5}$$

where k_p and k_i are the PI controller's gain coefficients. Equation (6)'s DC link loss component and fundamental load current component are added to get the total reference current magnitude (I_{tot}).

$$I_{tot} = I_{loss} + I_f \tag{6}$$

Now, the following equation is used to produce the unit template (u_a).

$$u_a = \frac{v_\alpha}{\sqrt{v_\alpha{}^2 + v_\beta{}^2}} \tag{7}$$

Multiplying the total reference current by the unit template yields the reference current.

$$I_{ref} = I_{tot} * u_a \tag{8}$$

Finally, gating pulses are produced for GSC's IGBT switches (insulated gate bipolar transistors) using a hysteresis current controller (HCC). In Fig. (2), the full control system is shown.

B. Modified LLC Converter

The equivalent model of the Bidirectional LLC converter is shown in Fig. (4). V_{in} is the total input voltage, while the output voltage is V_0. The resonant tank is composed of the resonant inductor L_r, two split capacitors C_1 and C_2, and two clamping diodes D_1 and D_2. The magnetizing inductor L_m of the isolated transformer has a turn ratio of n.

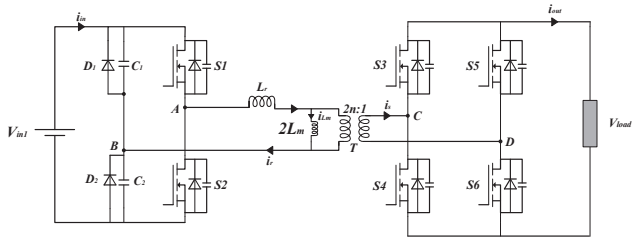


Fig. 4. Equivalent Bidirectional LLC resonant converter.

The energy stored in the junction capacitor must be smaller than that of the magnetizing inductor L_r in order to meet the ZVS requirements from the research.

$$\frac{1}{2} L_m i_{Lm_peak}^2 \geq \frac{1}{2}(C_{S3} + C_{S4} + C_{S5} + C_{S6}) V_0^2 . \tag{9}$$

where C_{S3}, C_{S4}, C_{S5}, and C_{S6} are the junction capacitors for the individual switches

where T_{dead} is dead time and is described below:

$$T_{dead} = \frac{1}{2 f_s} - \frac{1}{2 f_r} . \tag{10}$$

The range of L_m can be identified and expressed as follow, the range of the magnetizing inductance is,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:03 UTC from IEEE Xplore. Restrictions apply.

$$L_m \le \sqrt{\frac{n^2 V_0^2 L_r}{32 C_{S1} f_s V_{in1}}}. \quad (11)$$

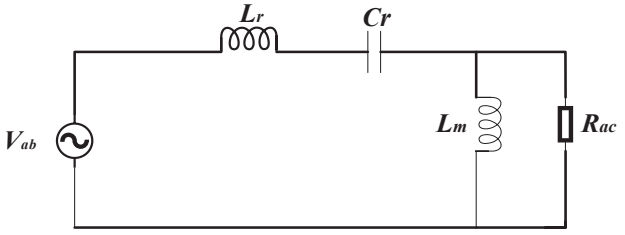


Fig. 5. Equivalent model of the LLC converter.

$$R_{ac} = \frac{8n^2 R_0}{\pi^2}. \quad (12)$$

The real load of the converter is R_0, and the resonating frequency is f_r.

$$f_r = \frac{1}{2\pi\sqrt{L_r C_r}}. \quad (13)$$

LLC converter Gain is described below:

$$M = \frac{f^2(m-1)}{\sqrt{(m.f^2-1)^2 + f^2(f^2-1)^2.(m-1)^2.Q^2}} \quad (14)$$

where

$$f = \frac{f_s}{f_r}. \quad (15)$$

$$Q = \frac{\sqrt{L_r / C_r}}{R_{ac}}. \quad (16)$$

$$m = \frac{L_r + L_m}{L_r}. \quad (17)$$

$$f_r = \frac{1}{2\pi\sqrt{L_r C_r}}. \quad (18)$$

Transformer turns ratio

The transformer turns ratio for this converter is described as:

$$n = \frac{V_{in}}{8V_0}. \quad (19)$$

Design of Resonant Tank

$$C_1 = \frac{P_{\max}}{2 f_s V_{in1}^2}. \quad (20)$$

where P_{max} is the maximum output power that each LLC converter is capable of producing. Additionally, L_r can be computed using,

$$Lr = \frac{1}{8\pi^2 C_1 f_s^2}. \quad (21)$$

C. PI Controller For DC-DC Converter

To determine an error signal, the reference current from the battery (I_{bref}) and the actual battery current (I_{bat}) are juxtaposed. The error signal is input into the PI controller, as shown in Fig. (1), and the PI controller's output is routed into the PWM block, which produces the electric throb for the isolated LLC converter. The Ziegler-Nichols frequency response technique is used to adjust the PI controller. The use of this technique requires a closed-loop PID controller. To acquire anticipated process features such as it is necessary to drive the controller to its stability boundaries. When there is a lesser dead time, this approach performs smoothly.

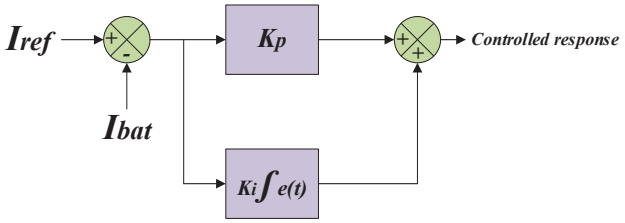


Fig. 6. PI controller for DC-DC converter stage.

IV. Simulation Results

By modeling under different operating circumstances, the efficacy of the proposed battery charger is verified. Additionally, the input to the battery charger is regarded to be an AC power source with 230 V RMS voltage and 50 Hz frequency. The designed charger's responsiveness is confirmed while variable grid voltage operations are obtained for three separate situations, taking into account the various operating circumstances. Fig. (7) shows the normal operation of the grid, Fig. (9) shows a non-ideal grid in which harmonics are introduced known as weak grid operation and Fig. (11) shows the 10% sag & swell in the grid voltage which is also a non-ideal case. The total behaviour of all three cases is shown and changes is seen in power factor, battery current, and SoC. The THD is calculated for every case and a comparison is done.

Particularly noteworthy are the simulation findings during the G2V operation under rated conditions shown in Figs. (7),(9),(11), and THD responses in Figs. (8),(10),(12). The waveforms of in-phase grid voltage, V_g, and grid current, I_g, illustrated in Fig. (7), indicate unity-power-factor (UPF) charging. Furthermore, the stabilized DC link voltage, V_{dc}, validates the VSC's efficacy. Additionally, I_{bat}'s controlled charging current demonstrates that the created charger efficiently controls the charging activity. Also, the battery current ripples adhere to the pertinent international standard. Fig. (8) also shows the fast Fourier-transform (FFT) evaluation of the grid current. The total harmonics distortion (THD) of grid current is indicated as 1.61% here. As a result,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:03 UTC from IEEE Xplore. Restrictions apply.

the designed charger's G2V functioning adheres to the power quality requirement, and the battery is being charged.

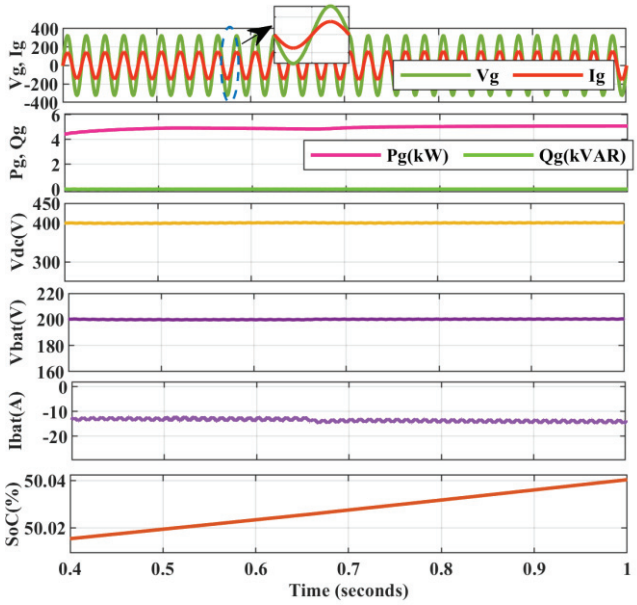


Fig. 7. Steady state performance under normal operating condition.

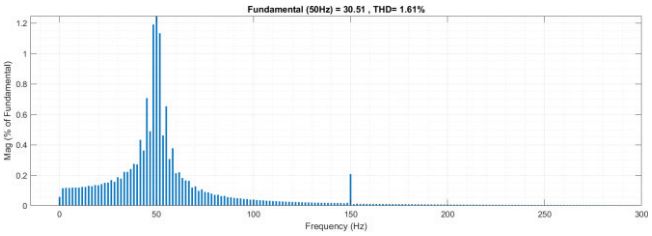


Fig. 8. THD under normal operating condition.

Fig. (9) depicts the system's weak grid operation, in which odd-ordered harmonics are added to grid voltage to test the system's robustness. As illustrated in Fig. (9), odd-ordered harmonics such as the 5^{th}, 7^{th}, 11^{th}, and 13^{th} distort

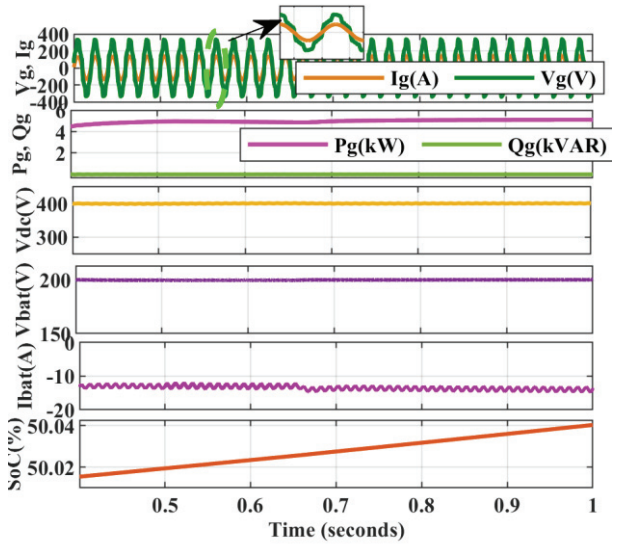


Fig. 9. Steady state performance when harmonics are injected (weak grid operating condition).

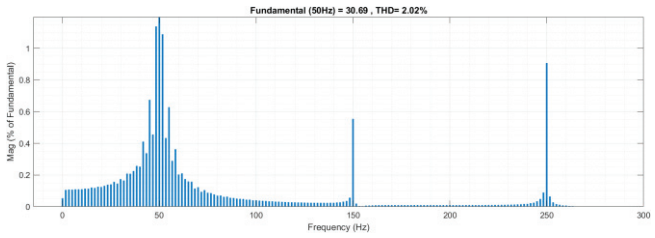


Fig. 10. THD under weak grid operation.

the grid voltage, but an input filter with SOGI control is far more efficient, attempting to align grid current with grid voltage to preserve power factor. V_{dc} and the battery charging current are both constant during this procedure. The predicted THD for this operation is 2.02%, which is likewise following the IEC requirement for grid current, as shown in Fig. (10).

Fig. (11) depicts the sag and swell of grid voltage under non-ideal conditions. During this operation, 10% of grid voltage is reduced between 0.5s and 0.6s, and 10% of grid voltage is increased between 0.7s and 0.8s, demonstrating the system's durability as V_{dc} is maintained constant throughout the operation, battery charging current is also kept constant.

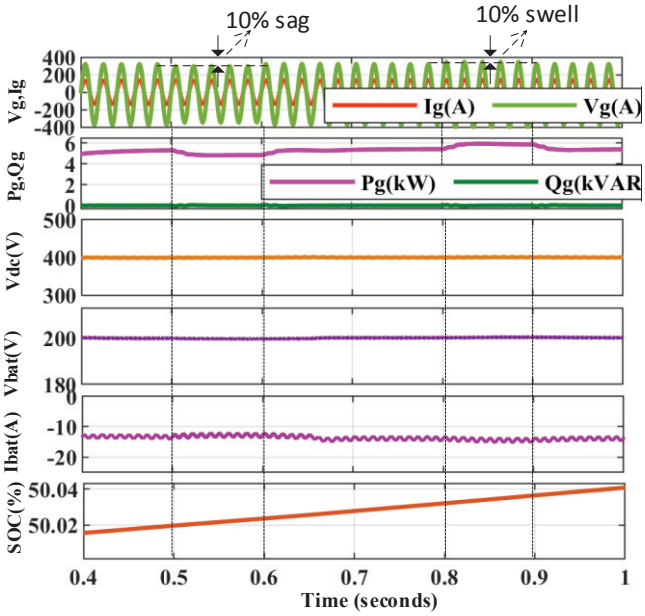


Fig. 11. Steady state performance when sag and swell in grid voltage.

Notably, THD adheres to the IEC norm for grid current quality, which is *1.87%* in this situation, as seen in Fig. (12).

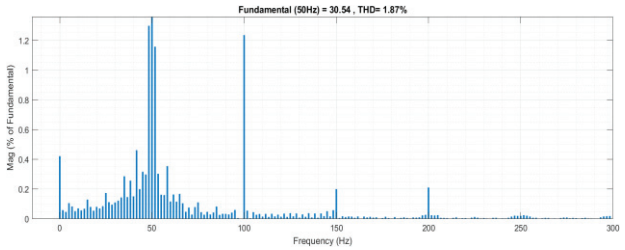


Fig. 12. THD under sag-swell voltage variation.

The resonating behaviour of the regulated MOSFET switches is seen in Fig. (13). To boost system efficiency with significantly reduced losses in the DC-DC conversion stage, all switches achieve ZVS at all conditions during G2V operating mode. All of the switches dissipate close to zero

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:03 UTC from IEEE Xplore. Restrictions apply.

power for the Modified LLC converter, which justifies the need for isolated converters in the system nicely.

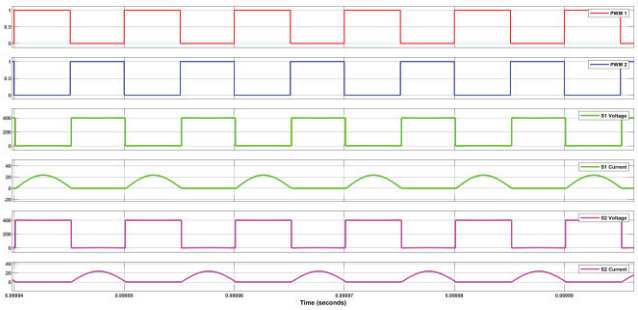


Fig. 13. Resonating nature of switches showing voltage and current with respect to the PWM.

V. CONCLUSION

This study investigates a two-stage On-Board charging system for different practical approaches. Two converters are introduced with a controlled scheme separately. AC-DC converter is controlled through a SOGI controller whereas a modified DC-DC converter is controlled through a PI controller and optimum results are obtained for diverse conditions. The functionality of the suggested technique is demonstrated by Matlab/Simulink results. Grid to Vehicle charging mode is designed for onboard EV chargers. At a frequency switching of 100 kHz, the onboard charger's planned system architecture was realized. The final findings demonstrate that the specified power efficiency is greater than 96%, the PF is over 99%, and the THD of grid current is less than 2% for the three separate situations described above. To demonstrate the reliability and efficiency of the SOGI controller for AC-DC converters and the PI controller for modified DC-DC converters, simulation studies are performed as well under steady state and dynamic conditions. For large voltage, large power EV charging applications, the LLC single phase On-Board charger offers inexpensive pricing, minimal electromagnetic interference, excellent dependability, compactness, and highly effective development. To assure converter performance within a wide voltage range for EV charging applications, a modified LLC converter's resonant behaviour is discovered inside the operational region using a second-order generalized integrated approach. Future experimental testing will be conducted to obtain a more precise picture of the overall effectiveness of your suggested system. For the DC-DC converter, a more durable and effective controller will be used to stabilize the dynamic behaviour of the battery side.

REFERENCES

[1] S. Chakraborty, H.-N. Vu, M. M. Hasan, D.-D. Tran, M. E. Baghdadi and O. Hegazy, "DC-DC Converter Topologies for Electric Vehicles, Plug-in Hybrid Electric Vehicles and Fast Charging Stations: State of the Art and Future Trends," Energies, vol. 12, no. 8, pp. 1-43, April 2019.

[2] M. Xiaochen, W. Ping, B. Hukan and W. Zhishuang, "A Bidirectional LLCL Resonant DC-DC Converter With Reduced Resonant Tank Currents and Reduced Voltage Stress of the Resonant Capacitor," IEEE Access, vol. 8, pp. 125549-125564, 2020.

[3] A. K. Seth and M. Singh, "Unified adaptive neuro-fuzzy inference system control for OFF board electric vehicle charger," International Journal of Electrical Power & Energy Systems, vol. 130, no. 0142-0615, p. 106896, 2021.

[4] K. M. Rizwan, I. A. Khan, H. Salman, A. M. S. Jamil and R. Jong-Suk, "A Comprehensive Review on Structural Topologies, Power Levels, Energy Storage Systems, and Standards for Electric Vehicle Charging Stations and Their Impacts on Grid," IEEE Access, vol. 9, pp. 128069-128094, 2021.

[5] B. Singh, B. N. Singh, A. Chandra, K. Al-Haddad, A. Pandey and D. P. Kothari, "A review of three-phase improved power quality AC-DC converters," IEEE Transactions on Industrial Electronics, vol. 51, no. 3, pp. 641-660, 2004.

[6] H. Yap, R. M. A. Mohd, H. M. Khair and M. N. Farzilah, "Operation of Three-Level Inverter-Based Shunt Active Power Filter Under Nonideal Grid Voltage Conditions With Dual Fundamental Component Extraction," IEEE Transactions on Power Electronics, vol. 33, no. 9, pp. 7558-7570, 2018.

[7] K. Parag, B. Singh, C. Ambrish and A.-H. Kamal, "SRF Theory Revisited" to Control Self-Supported Dynamic Voltage Restorer (DVR) for Unbalanced and Nonlinear Loads," IEEE Transactions on Industry Applications, vol. 49, no. 5, pp. 2330-2340, 2013.

[8] W. Fengjiang, Z. Lujie and D. Jiandong, "A New Two-Phase Stationary-Frame-Based Enhanced PLL for Three-Phase Grid Synchronization," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 62, no. 3, pp. 251-255, 2015.

[9] G. Saeed, G. M. Josep, M. Fariborz and V. J. C, "Single-Phase Frequency-Locked Loops: A Comprehensive Review," IEEE Transactions on Power Electronics, vol. 34, no. 12, pp. 11791-11812, 2019.

[10] H. Nie and L. Y. Wei, "Overview and Comparison of Modulation and Control Strategies for a Nonresonant Single-Phase Dual-Active-Bridge DC–DC Converter," IEEE Transactions on Power Electronics, vol. 35, no. 3, pp. 3148-3172, 2020.

[11] S. Dongdong and W. Haoyu, "Light-Load Performance Enhancement Technique for LLC-Based PEV Charger Through Circuit Reconfiguration," IEEE Transactions on Transportation Electrification, vol. 7, no. 4, pp. 2104-2113, 2021.

[12] K. N. Zare, P. M. Aldi, E. Mojtaba, B. A. A. Moti and A. Parviz, "Grid Synchronization of Bidirectional Electric Vehicle Chargers Using Second Order Generalized Integrator based Phase Lock Loop," in 2021 12th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC), 2021, pp. 1-5.

[13] K. Aswini, J. Kamala, L. Sriram, B. Kowshik, B. R. Vara Prasad and D. V. Sai Bharani, "Design and Analysis of Bidirectional Battery Charger for Electric Vehicle," INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT), vol. 10, no. 7, 2021.

[14] S. Wenjin, X. Yan, W. Hongfei and D. Jie, "Modified High-Efficiency LLC Converters With Two Split Resonant Branches for Wide Input-Voltage Range Applications," IEEE Transactions on Power Electronics, vol. 33, no. 9, pp. 7867-7879, 2018.

[15] S. Tong, R. Xiaoyong, C. Qianhong, Z. Zhiliang and R. Xinbo, "Reliability and efficiency improvement in LLC resonant converter by adopting GaN transistor," in 2015 IEEE Applied Power Electronics Conference and Exposition (APEC), 2015, pp. 2459-2463.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:03 UTC from IEEE Xplore. Restrictions apply.

Preprints.org

Article

Not peer-reviewed version

A Compact Wideband SRR-based C-Shaped Antenna for 5G New Radio Band n258

Suresh Angadi , Thennarasan Sabapathy * , N.S. Raghava , Muzammil Jusoh , Hamsakutty Vettikalladi , Ali M. Almuhlafi , Mohamed Himdi

Posted Date: 11 August 2023

doi: 10.20944/preprints202308.0963.v1

Keywords: SRR; millimeter wave; 5G; Wide-band Antenna

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Compact Wideband SRR-Based C-Shaped Antenna for 5G New Radio Band n258

Suresh Angadi [1], Thennarasan Sabapathy [2,3,*], N.S. Raghava [1], Muzammil Jusoh [2,3], Hamsakutty Vettikalladi [4], Ali M. Almuhlafi [4] and Mohamed Himdi [5]

[1] Department of E.C.E, Delhi Technological University (DTU), Delhi, India-110042; suresh.angadi.101@gmail.com; nsraghava@gmail.com;
[2] Advanced Communication Engineering (ACE), Centre of Excellence, Universiti Malaysia Perlis (UniMAP), Jalan Tiga, Pengkalan Jaya Business Centre, Kangar 01000, Perlis, Malaysia; muzammil@unimap.edu.my.com
[3] Faculty of Electronic Engineering Technology, Universiti Malaysia Perlis (UniMAP), Kampus Alam UniMAP Pauh Putra, Arau 02600, Perlis, Malaysia;
[4] Electrical Engineering Department, College of Engineering, King Saud University, Riyadh 11421, Saudi Arabia; hvettikalladi@ksu.edu.sa; aalbishi@ksu.edu.sa
[5] Institut d'Electronique et des Technologies du numéRique (IETR), University of Rennes, 35042 Renne, France; mohamed.himdi@univ-rennes1.fr
* Correspondence: thennarasan@unimap.edu.my

Abstract: This paper proposes a compact wideband split ring resonator (SRR) based C-shaped antenna for 5G new radio (NR) band n258 millimeter wave (mm-Wave). The proposed antenna covers a wideband frequency range from 24.1 GHz to 28.3 GHz with a reflection coefficient below -10 dB. Stubs are used at each C-shaped element for impedance matching of the antenna. The proposed antenna comprises SRRs with stubs to increase the bandwidth and gain simultaneously for better performance in mm-Wave applications. A simple SRR structure is modeled to function as a reactive impedance surface with inductive property for frequencies less than 30.2 GHz. The proposed antenna realizes an impedance bandwidth of 16.1% and covers a wideband of 4.2 GHz (24.1 GHz -28.3 GHz). The use of stubs and SRR has improved the performance of the proposed antenna at mm-Wave frequencies. Specifically, the deployment of the SRR is able to produce an improved gain for the overall direction of the antenna. The simulation and measured results agree well and maintain a 7.49 dBi gain with a bandwidth of 4.2 GHz. An average efficiency of 90% can be attained in the operating range of the antenna. The designed antenna is applicable for 5G NR Band n258.

Keywords: SRR; millimeter wave; 5G; wide-band antenna

1. Introduction

The viability of 5G networks extends the facility of spectrum utilization, network flexibility, and multi-connectivity in the sphere of wireless technology. The development of 5G increases the high data rate and improves the occupancy of bandwidth in many communication devices, allowing higher mobile users per unit area [1]. In 5G millimeter wave technology, various approaches have been taken to enhance bandwidth, improve antenna gain, and achieve miniaturization. These approaches could help invent future innovations that meet millimeter wireless standards and play a key role in the communication medium.

To realize a wide bandwidth in millimeter wave frequencies, a defected ground approach [2] was used in antenna design. A new wideband array antenna was considered for mm-Wave applications intended for high data rates in wireless communications [3]. In other related work, inverted L-slots were placed at radiating elements to improve the bandwidth of an individual feed and match the impedance parameter of the antenna for millimeter wave applications in 5G mobile

© 2023 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

applications [4]. An electromagnetic band gap (EBG) ground was positioned side by side in the antenna array for better matching and improved radiating properties of the antenna, making it acceptable for enabling mobile terminals in 5G communication [5]. Slots can be added to improve the bandwidth in mm-Wave frequencies and reduce the effective area in the radiating patch [6]. An EBG-based substrate-integrated waveguide (SIW) antenna has been shown and proven to operate in mm-Wave applications [7].

Apart from bandwidth enhancement, gain improvement is also considered one of the key achievements in 5G antenna designs. The split ring resonator (SRR) has been introduced in antenna design to improve gain in 5G wireless technology that operates at the low band (600-850 MHz), mid-band (3.4-3.6 GHz), and higher band in millimeter wave (mm-Wave) frequencies of (24-28) GHz, as reported in [8-9]. Array antennas have also been used for mm-Wave 5G smartphones to improve gain in the antenna design, as reported in [10]. In addition, dielectric resonator antennas (DRA) have been used to improve gain and bandwidth for mm-Wave applications, as reported in [11-12]. In the view of wireless communication under 5G technology, a dual E-shaped patch antenna generated a multiband with stubs inserted between the elements and enhanced the wideband property into ultra-wideband, which has shown better results in terms of gain, as reported in [13,14]. Furthermore, meta-surface based approach has also been used in 5G antennas to improve gain, as reported in [15].

In addition to bandwidth and gain enhancement, the design of 5G antennas is also considered in terms of miniaturization. A basic square-shaped patch loaded with a slot was proposed in [16] to achieve miniaturization. An optimized low-profile antenna is designed using an artificial magnetic conductor reflector, maintaining low cross-polarization and increasing gain and bandwidth with antenna efficiency [17]. A miniaturized mm-Wave antenna has also been designed for a multiple-input multiple-output 5G antenna that operates with beam scanning ability [18]. Nevertheless, all reported works still lack in achieving a compact antenna that operates at mm-Wave 5G new radio (NR) band and attains high gain for an omnidirectional radiation pattern.

In this work, a C-shaped patch antenna was modeled with simple split ring resonators (SRR) to improve the quasi-omnidirectional radiation pattern over a wideband for 5G mm-Wave applications. Rectangular stubs were used preceding every portion of the C-shaped patch to eliminate the dual band and enable the wideband property of the antenna. In general, monopole microstrip patch antennas face limitations such as narrow bandwidth [19] and low gain [20]. To overcome these limitations, careful analyses were carried out, and the proposed design was optimized and investigated at each phase to finally develop the antenna with wide-band, omnidirectional pattern, and high gain characteristics. A compact wideband antenna was proposed in frequency range 2 (FR2) 5G NR band n258. The impedance bandwidth of 16.1% is achieved, covering two bands centered at 25.2 GHz and 27.7 GHz frequencies.

2. Antenna Design and Analysis

The geometry of the compact wideband SRR-based antenna for 5G mm-Wave application is presented in **Error! Reference source not found.**. The designed wideband antenna was printed on a Rogers RT/duroid 5880 substrate with a relative permittivity of 2.2, and a dielectric loss tangent of 0.0009. The overall dimension of the antenna is covered by the length L_s, width W_s, and substrate height of 1.6 mm. S_L and S_W are the length and width of the stub. **Error! Reference source not found.** shows the complete dimensions of the proposed antenna. The dimensions of the ground plane G_p are placed at the lowermost sheet of the antenna to enable omni-directional radiation pattern characteristics. At each element in the array, the rectangular stubs are used prior to the C-shaped structure to reduce the losses and widen the bandwidth of the antenna as shown in the parametric analysis. Then, an SRR is used to improve the radiation pattern characteristic of the antenna. Finally, 4 elements of C-shaped radiating element SRR could be developed as shown in Figure 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

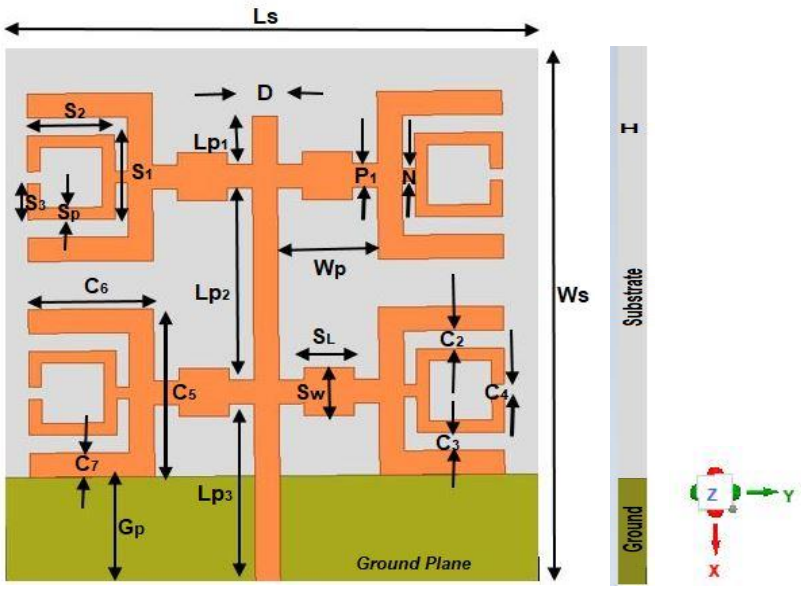


Figure 1. The Geometry of SRR-based C shaped antenna.

Table 1. Parametric Dimension table for the proposed antenna.

Dimension	Value	Dimension	Value	Dimension	Value
L_S	43	D	2	C_7	2
W_S	45	S_L	4	S_1	7
G_P	9	S_W	4	S_2	7
H	1.6	W_P	8	S_3	6.5
L_{P1}	4	P_1	2	S_P	1
L_{P2}	16	C_4	1	N	1
L_{P3}	15	C_5	14		
C_3	1.8	C_6	10		

[1] All Dimensions are in mm.

2.1. SRR unit cell design

Figure 2 (a) shows the SRR unit cell design along with its reflection phase result. A simple 7 mm x 7 mm square-shaped SRR unit cell was adopted with a ring width of 1 mm. The slit width between the ring is C_4. The boundaries of the SRR unit cell were covered by a perfect electric conductor (PEC) and perfect magnetic conductor (PMC) perpendicular to the incident electric (E) and magnetic (H) fields. This SRR was subjected to transverse electromagnetic waves (TEM) coming from the +z direction. As shown in Figure 2(b), the SRR acts as an inductor below the center frequency at 30.2 GHz which can help to store the magnetic energy that could compensate for the electrical energy [21]. This reflection phase characteristic of the SRR is expected to enhance the radiation characteristics of the antenna within the operating range from 24.2 to 28.3 GHz.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

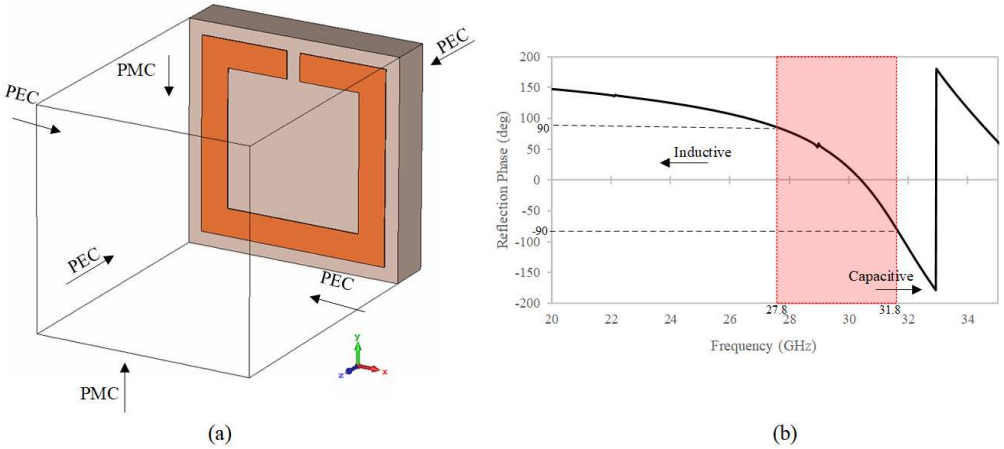


Figure 2. The SRR design: (a) unit cell (b) Reflection phase result.

2.2. Equivalent circuit model of the SRR-based C-Shaped antenna

The equivalent circuit for the compact wideband SRR-based antenna is shown in Figure 3. Here the circuit is established based on resistor (*R*), inductor (*L*), and capacitor (*C*), and the equivalent circuit is considered only for one element in the array from the feed line. Z_c is the characteristic impedance and R_f is the loss present in the feed line. The equivalent circuit model is adopted based on [22] and modified according to the dimension of the proposed antenna in this work. The inductor L_1 and Capacitor C_1 are represented [22] as microstrip lines in the equivalent circuit.

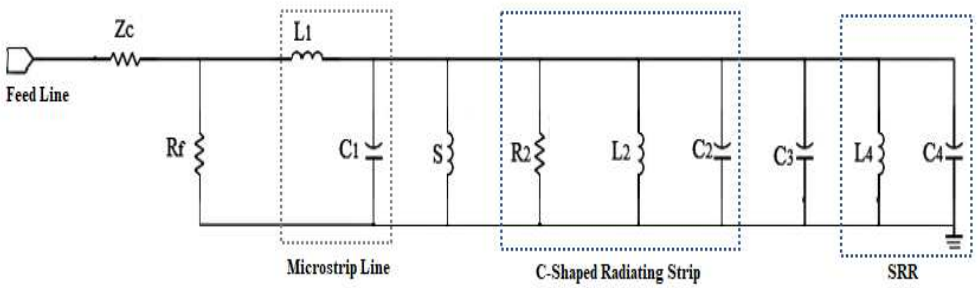


Figure 3. Equivalent Circuit Model for a single element in the proposed antenna.

The mathematical expression for L_1 and C_1 are represented as follows.

$$L_1 = 100H\left(4\sqrt{\frac{D}{H} - 4.21}\right)nH \tag{1}$$

$$C_1 = D\left\{(9.5E_r + 1.25)\frac{D}{H} + 5.2E_r + 7\right\}pF \tag{2}$$

where *D* and *H* represent the width of the feedline and height of the substrate respectively and E_r is the relative permittivity. In the equivalent circuit model, *S* represents a stub placed for impedance matching and for widening the bandwidth. The width (W_{stub}) and length (L_{stub}) of the stub [13] are expressed in equations (3) and (4) respectively.

$$L_{stub} = \frac{1}{f_r}\sqrt{\frac{c}{\frac{E_r + 1}{2}}} \tag{3}$$

$$W_{stub} = S_w + D \tag{4}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

where c is the speed of the light and f_r is the resonant frequency. Based on the C_5 and C_6 dimensions as shown in Figure 1, the effective length (L_{eff1}) and width (W_{eff1}) of the C-Shaped radiating strip for one element can be calculated as follows,

$$L_{eff1} = C_5 + C_6 \tag{5}$$

$$W_{eff1} = C_7 \tag{6}$$

The expression for the capacitor (C_2), resistor (R_2), and inductor (L_2) of the equivalent circuit can be expressed in equations 7, 8, and 9.

$$C_2 = \frac{L_{eff1} W_{eff1} E_{eff}}{2H} \cos^2 \frac{\pi \left(L_{p3} + G_p \right)}{L_{eff1}} \tag{7}$$

$$R_2 = \frac{Q}{\left(2\pi f_r\right)^2 C_2} \tag{8}$$

$$L_2 = \frac{1}{\left(2\pi f_r\right)^2 C_2} \tag{9}$$

where the quality factor, Q can be expressed as,

$$Q = \frac{c\sqrt{E_{eff}}}{4 f_r} \tag{10}$$

In the equivalent circuit model, L_4 is the inductance and C_4 is the capacitance for the square split ring resonator. For a square split ring resonator, the resonant frequency f_{srr} is expressed as

$$f_{srr} = \frac{1}{2\pi} \sqrt{\frac{1}{L_4 + C_4}} \tag{11}$$

where C_4 is the total equivalent capacitance of the SRR,

$$C_4 = \frac{L_4 \sqrt{E_{eff}}}{c Z_c} + \frac{E_O N H}{2 L_4} \tag{12}$$

where Z_C is the characteristic impedance and L_4 is the total equivalent inductance of the SRR. The L_4 can be expressed as follows,

$$L_4 = 0.0002 L_{eff1} \left(2.303 \log_{10} \frac{4 L_{eff1}}{c} - \gamma \right) \mu H \tag{13}$$

where γ = 2.853 for a wire loop of the square geometry and the L_{eff1} can be calculated based on the total length of the SRR as,

$$L_{eff1} = s_1 + s_2 + s_3 \tag{14}$$

3. Results and Discussion

The simulation was carried out using the HFSS EM modeling tool. The proposed antenna was fabricated and tested using the Vector Network Analyzer of R & S ZND 40. The S-parameter and peak gain results were observed in each simulation stage to observe the effect of the sub-matching and the SRR.

3.1. The Design Evolution of The Proposed Antenna

The evolution process of the proposed antenna is shown in **Error! Reference source not found.Error! Reference source not found.**. The reflection coefficient response for each stage is shown

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

in Figure 4. of the proposed SRR-based antenna. Antenna 1 with diameter D = 2mm resonates with two resonant frequencies at 25.4 and 27.8 with S_{11} of -24.07 and -14.71 and covers a bandwidth in each band at 1.2 and 0.9 with a gain of 9.22. The addition of two stubs in antenna 2 increases the bandwidth of the antenna and resonates at two frequencies with an increase in S_{11} of -36.05 at 25.5 and -16.06 at 27.8 with a bandwidth of 1 GHz and 1.1 GHz. In antenna 3, the stubs were placed before each element of the antenna and it covers the wider bandwidth of 4.1 GHz and resonates at two frequencies of 25.1 and 27.8 GHz with a reflection coefficient of -16.6 and -15.29, and attains a peak gain of 7.1 dB. The proposed Array Antenna resonates at 25.2 and 27.7 with S11 of -18.43 and -16.11 with an increased peak gain of 7.49 dB as shown in Figure 5(b). The results from the design evaluation are summarized in **Error! Reference source not found.** 3.

Table 2. The evolution process of SRR based C-Shaped Antenna.

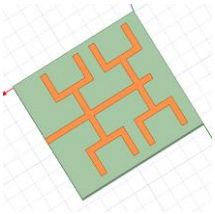
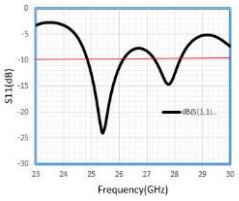
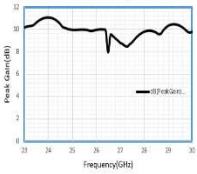
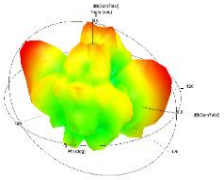
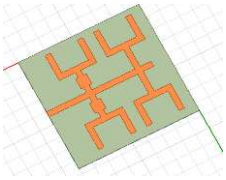
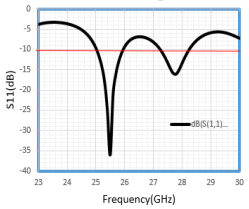
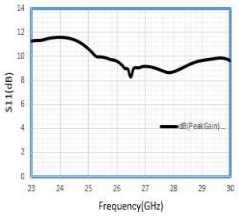
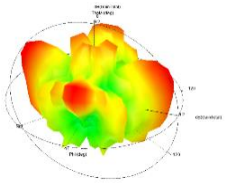
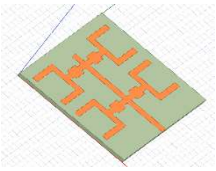
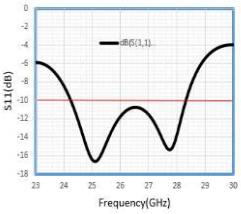
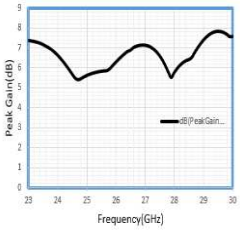
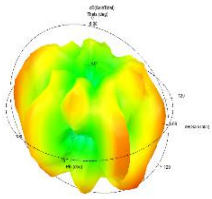
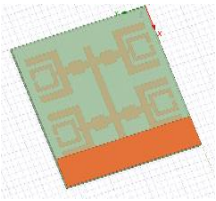
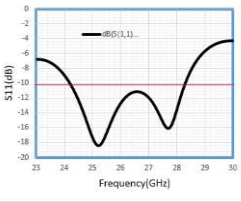
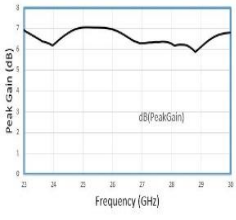
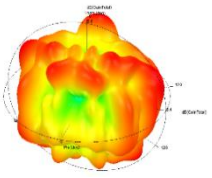
Antenna Evolutions	3D Model	S-Parameter	Gain Plot	3D Plot
Antenna 1				
Antenna 2				
Antenna 3				
Proposed Antenna				

Table 3. Summary results of design evaluation.

Antenna Steps	Operating Frequency(GHz)	S_{11} (dB)	B.W(GHz)	Peak Gain (dB)
Antenna-1	25.4	-24.07	1.2	9.22
	27.8	-14.71	0.9	
Antenna-2	25.5	-36.05	1	8.3044
	27.8	-16.06	1.1	
Antenna-3	25.1	-16.6	4.1	7.1
	27.8	-15.29		

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

Proposed	25.2	-18.43		
Antenna	27.7	*-16.11*	4.2	7.49

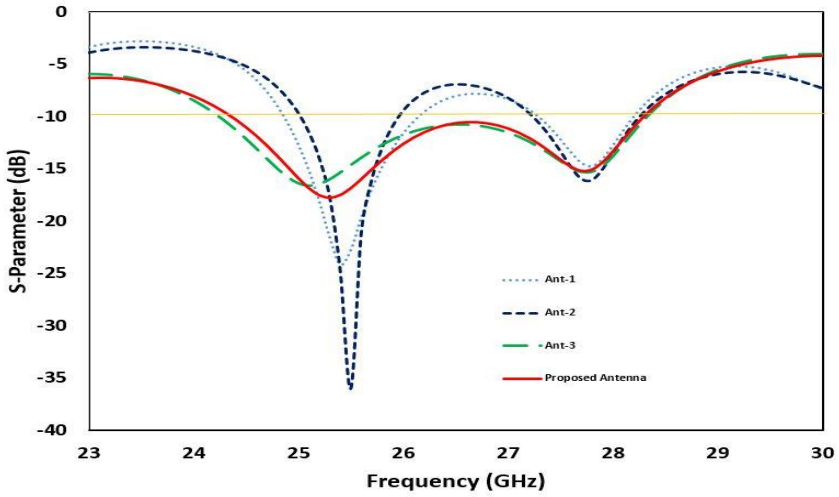
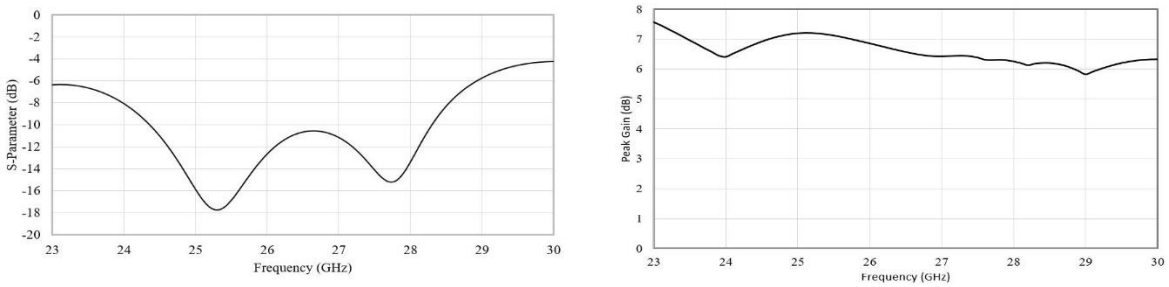


Figure 4. Parametric Analysis SRR-based C-Shaped Antenna.



(**a**) (**b**)

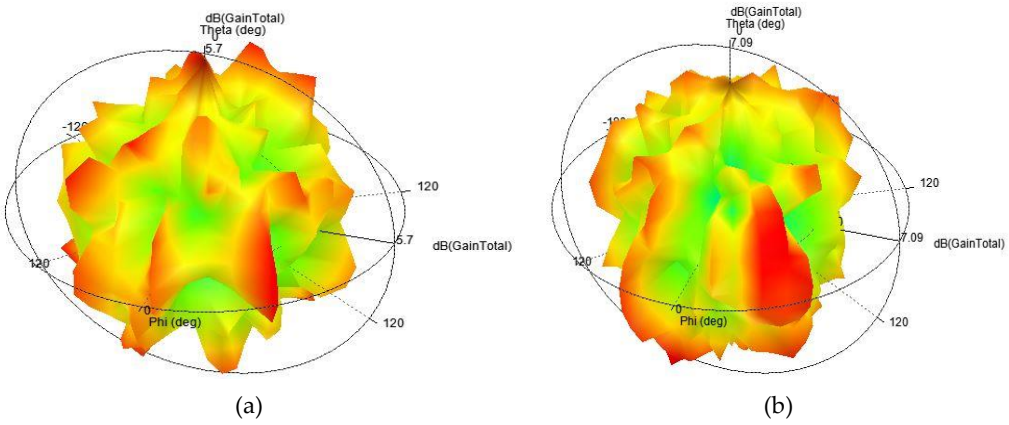
Figure 5. Simulation results of SRR-based C-Shaped Antenna: (a) Response of S11 Vs Frequency; (b) Response of Peak Gain Vs Frequency.

The use of SRR to improve the overall gain of the antenna may not be clearly observed through this 2D radiation pattern result, because it does not show the radiation pattern characteristic for the overall beam coverage. Therefore, in what follows, the 3D radiation pattern of the antenna is investigated.

3.2. D Radiation Pattern Analysis

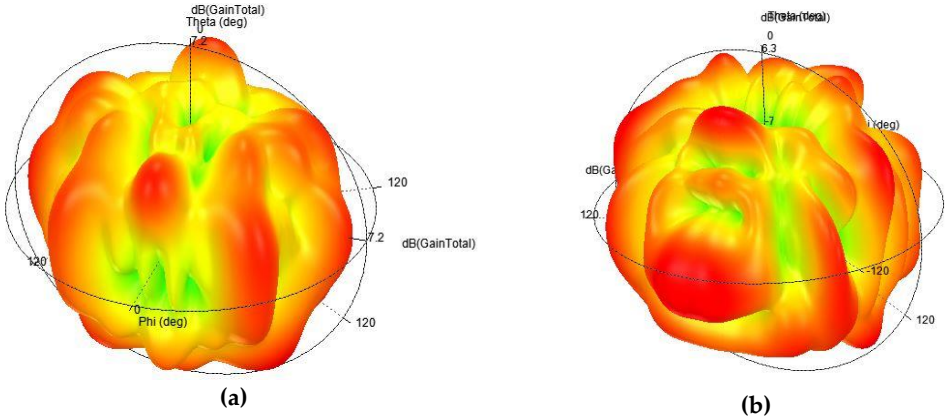
In this design, the SRR mainly helps to improve the antenna gain for most of the directions, and thus attains an omnidirectional radiation pattern. From Table 2. It can be seen that the antenna's reflection coefficient remains the same for Antenna 3 and the proposed antenna. This explains that the wide-band property was obtained mainly from the use of stubs. The 2D radiation patterns shown in Table 2 are also almost similar for both Antenna 3 and the proposed antenna, just that a slight improvement can be obtained in terms of the peak gain. To further observe the advantage of the SRR, the 3D radiation pattern results are compared. Figure 6 and Figure 7 show the radiation pattern results for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1



(a) (b)

Figure 6. 3D radiation pattern results for Antenna 3 at :(a) 25.1 GHz and; (b) 27.8 GHz.



(a) **(b)**

Figure 7. 3D radiation pattern results for the proposed antenna at :(a) 25.1 GHz and; (b) 27.8 GHz.

Antenna 3 and the proposed antenna, respectively. The results also compared two operating frequencies, 25.1 GHz and 27.8 GHz. Compared with Antenna 3, the proposed antenna manages to radiate with more gain for a wide range of angles. Hence, it shows that the high gain is not limited to certain directions/angles but covers most of the directions with the use of the SRR. The results were also further validated with the observation of the total efficiency and the peak gain proposed antenna and the antenna without SRR. It can be seen from Figure 8 that the antenna efficiency reaches 99% with better gain in the operating range of the antenna.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

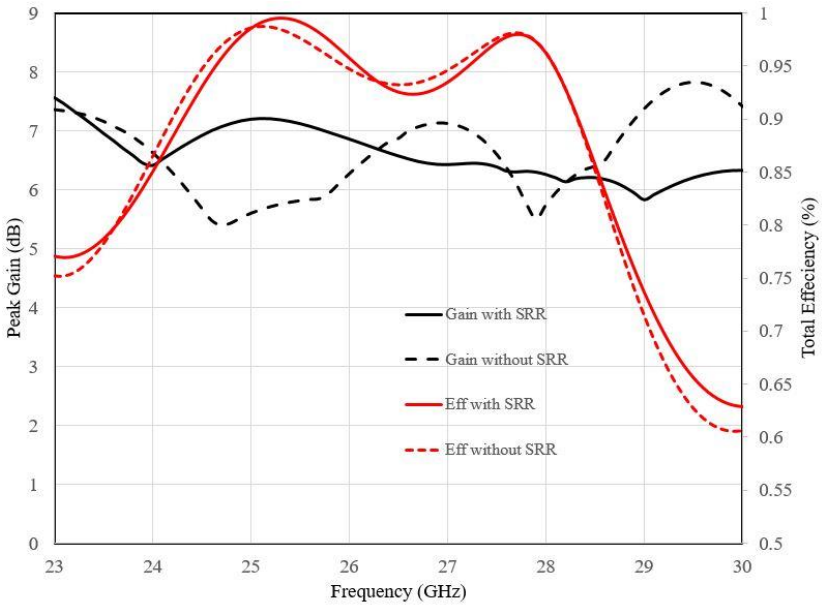


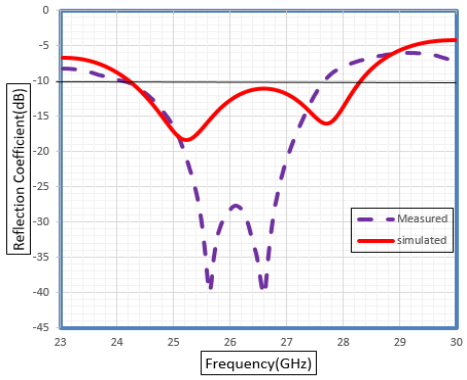
Figure 8. Total efficiency and peak gain of the proposed Antenna.

3.3. The Fabricated Antenna and The Measurement Results

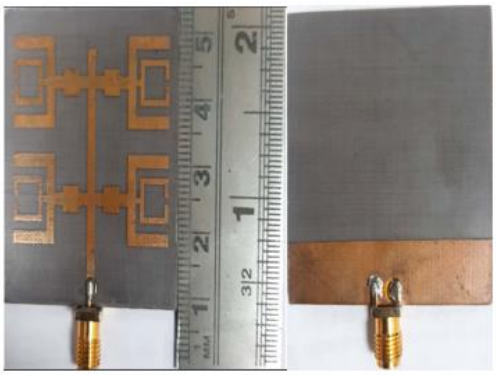
The measurement and simulation results of S_{11} against frequency are shown in Figure 9 (a). The top view and rear view of the fabricated Antenna are shown in Figure 9 (b). The measurement was made in the laboratory using a Vector network analyzer (R & S ZND 40). After measuring the proposed wideband antenna resonates at 25.2GHz and 27. 7 GHz in the mm-Wave band (24.1 GHz - 28.3 GHz) below -10 dB. Figure 9. Fabrication and Measurement Results of the proposed antenna: (a) Simulation and Measurement results of the Proposed Antenna; (b) Fabrication of the proposed antenna; (c) Measurement Set-up; (d) VNA Measured Results.

shows the comparison of simulated and measured results relevant to the impedance bandwidth. The impedance bandwidth of simulated results is 16.1(%) and 13.40(%) is for measured results. This slight mismatch could be due to the soldering effect of the SMA port on the transmission line.

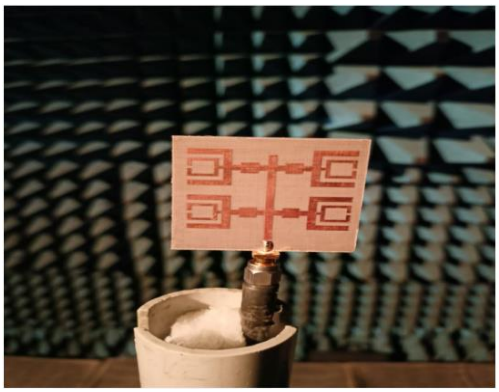
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1



(a)



(b)



(c)

(d)

Figure 9. Fabrication and Measurement Results of the proposed antenna: (a) Simulation and Measurement results of the Proposed Antenna; (b) Fabrication of the proposed antenna; (c) Measurement Set-up; (d) VNA Measured Results.

Table 4. Comparison Results of Proposed Antenna.

Type	Reflection Coefficient (dB)	Bandwidth (GHz)	Impedance Bandwidth (%)
Simulated	below -10dB	4.2	16.03
Measured	below -10dB	3.5	13.4

Figure 10 shows the polar radiation pattern cut xz plane for 25.2 and 27.7 GHz frequencies. It shows that a high gain can be obtained for most directions with acceptable cross-polarization.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

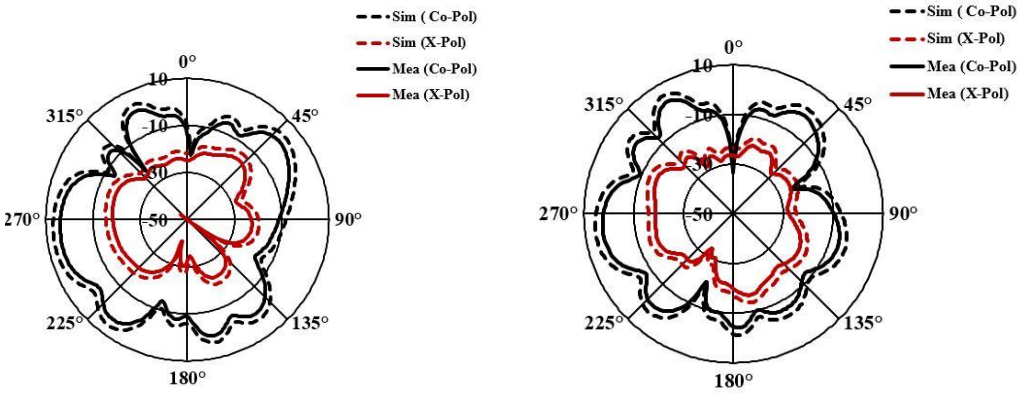


Figure 10. The polar radiation patterns at xz-plane: (a) 25.2 GHz and; (b) 27.7 GHz.

Table 5 presents the comparison of the proposed antenna with similarly proposed works for mm-Wave operating frequencies. In terms of dimension, the proposed antenna has a comparable size to most of the works. Only works [23] and [24] have smaller sizes compared with the proposed antenna. In terms of bandwidth and gain the proposed antenna have better results compared with most of the existing works reported in the literature. Specifically, the use of SRR cannot be compared with the existing works as the 3D radiation patterns are not revealed. Therefore, the advantage of the SRR was comprehensively analyzed and discussed.

Table 5. Comparison of the proposed antenna performance with the existing models.

Reference	Dimensions (LXW) $(mm)^2$	Coverage bands (GHz)	Band width (GHz)	Peak gain (dBi)
[7]	21x85	28	1.95	10.17
[11]	20x40	26.71-28.91	2.2	7
[16]	61x54	28	0.83	6.68
[18]	90x160	24.25-27.50	3.25	14
[23]	15x12.7	28	3	7
[24]	10x10	24.85	1.318	6.48
Proposed Antenna	**43x45**	**24.1-28.3**	**4.2**	**7.49**

4. Conclusion

This paper proposed a C-shaped SRR-based 5G mm-Wave antenna, which is a significant contribution to the field of wireless communication. The antenna's stub matching technique improves its bandwidth, while the SRR enhances its radiation characteristics. This study modeled a simple SRR to have inductive properties for frequencies below 30.2 GHz, which enables the antenna to have better radiation characteristics at its operating range. The proposed antenna operates with a wide band that resonates at 25.2 GHz and 27.7 GHz frequencies, with an impressive impedance bandwidth of 4.02 (24.1-28.3 GHz) and a reflection coefficient of -18.431 dB and -16.113 dB, respectively. The proposed compact wideband array antenna has an acceptable peak gain of 7.49 dBi with the use of SRR. These results outperform most of the existing works reported in the literature in terms of bandwidth and gain. Furthermore, the proposed antenna is applicable for use in 5G NR Band n258 FR2 (mm-Wave) at the 5G spectrum. The comprehensive analyses and discussions of the SRR's advantages in this paper demonstrate its potential to improve the antenna's radiation properties and bandwidth. Overall, the proposed C-shaped SRR-based 5G mm-Wave antenna is a promising candidate for future wireless communication systems and can contribute to the development of 5G technology.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

Supplementary Materials: Not applicable.

Author Contributions: Conceptualization, Suresh Angadi and Thennarasan Sabapathy; Data curation, Mohamed Himdi; Formal analysis, N.S. Raghava; Funding acquisition, Hamsakutty Vettikalladi and Ali M. Almuhlafi; Investigation, Suresh Angadi, Muzammil Jusoh and Hamsakutty Vettikalladi; Methodology, Suresh Angadi and N.S. Raghava; Project administration, Hamsakutty Vettikalladi; Resources, N.S. Raghava; Software, Ali M. Almuhlafi; Validation, Thennarasan Sabapathy and Ali M. Almuhlafi; Visualization, Muzammil Jusoh; Writing – original draft, Suresh Angadi and Thennarasan Sabapathy; Writing – review & editing, Thennarasan Sabapathy and Mohamed Himdi. All authors have read and agreed to the published version of the manuscript.

Funding: Researchers Supporting Project number (RSP2023R482), King Saud University, Riyadh, Saudi Arabia.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Patents: Not applicable.

References

1. J. R. James and P. S. Hall, Eds., *Handbook of Microstrip Antennas, Volume 1*. Institution of Engineering and Technology, 1989. doi: 10.1049/PBEW028F.
2. M. Khalid *et al.*, "4-Port MIMO Antenna with Defected Ground Structure for 5G Millimeter Wave Applications," *Electronics*, vol. 9, no. 1, p. 71, Jan. 2020, doi: 10.3390/electronics9010071.
3. H. Alwareth, I. M. Ibrahim, Z. Zakaria, A. J. A. Al-Gburi, S. Ahmed, and Z. A. Nasser, "A Wideband High-Gain Microstrip Array Antenna Integrated with Frequency-Selective Surface for Sub-6 GHz 5G Applications," *Micromachines*, vol. 13, no. 8, p. 1215, Jul. 2022, doi: 10.3390/mi13081215.
4. M. Sanad, "Comments on 'Wide-band E-shaped patch antennas for wireless communications,'" *IEEE Trans. Antennas Propagat.*, vol. 51, no. 9, p. 2541, Sep. 2003, doi: 10.1109/TAP.2002.807369.
5. E. M. Wissem, I. Sfar, L. Osman, and J.-M. Ribero, "A Textile EBG-Based Antenna for Future 5G-IoT Millimeter-Wave Applications," *Electronics*, vol. 10, no. 2, p. 154, Jan. 2021, doi: 10.3390/electronics10020154.
6. P. Merlin Teresa and G. Umamaheswari, "Compact Slotted Microstrip Antenna for 5G Applications Operating at 28 GHz," *IETE Journal of Research*, vol. 68, no. 5, pp. 3778–3785, Sep. 2022, doi: 10.1080/03772063.2020.1779620.
7. A. Verma and S. N. Raghava, "Circularly polarized hybrid mode substrate integrated waveguide antenna for two quadrant scanning beamforming applications for 5G," *Int J RF Microw Comput Aided Eng*, vol. 31, no. 10, Oct. 2021, doi: 10.1002/mmce.22798.
8. Wen Tao Li, Yong Qiang Hei, Wei Feng, and Xiao Wei Shi, "Planar Antenna for 3G/Bluetooth/WiMAX and UWB Applications With Dual Band-Notched Characteristics," *Antennas Wirel. Propag. Lett.*, vol. 11, pp. 61–64, 2012, doi: 10.1109/LAWP.2012.2183671.
9. T. S. Rappaport *et al.*, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," *IEEE Access*, vol. 1, pp. 335–349, 2013, doi: 10.1109/ACCESS.2013.2260813.
10. W. Hong, K.-H. Baek, and S. Ko, "Millimeter-Wave 5G Antennas for Smartphones: Overview and Experimental Demonstration," *IEEE Trans. Antennas Propagat.*, vol. 65, no. 12, pp. 6250–6261, Dec. 2017, doi: 10.1109/TAP.2017.2740963.
11. N. S. Murthy, "Improved Isolation Metamaterial Inspired Mm-Wave MIMO Dielectric Resonator Antenna For 5g Application," *PIER C*, vol. 100, pp. 247–261, 2020, doi: 10.2528/PIERC19112603.
12. M. D. Alanazi, "A Review of Dielectric Resonator Antenna at Mm-Wave Band," *Eng*, vol. 4, no. 1, pp. 843–856, Mar. 2023, doi: 10.3390/eng4010051.
13. K. K. Naik, "Improvement of wider bandwidth using dual e–shaped antenna for wireless communications in 5G applications," *Analog Integr Circ Sig Process*, vol. 109, no. 1, pp. 93–101, Oct. 2021, doi: 10.1007/s10470-021-01919-7.
14. "Wideband Slotted Planar Inverted-F Antenna using Eccosorb MCS Absorber for Millimeter-Wave Applications," *IJPAP*, Jun. 2021, doi: 10.56042/ijpap.v59i6.47802.
15. S. Tariq, S. I. Naqvi, N. Hussain, and Y. Amin, "A Metasurface-Based MIMO Antenna for 5G Millimeter-Wave Applications," *IEEE Access*, vol. 9, pp. 51805–51817, 2021, doi: 10.1109/ACCESS.2021.3069185.
16. M. Ikram, M. S. Sharawi, and A. Shamim, "A millimeter-wave connected antenna array for 5G applications," in *2017 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, San Diego, CA, USA: IEEE, Jul. 2017, pp. 1449–1450. doi: 10.1109/APUSNCURSINRSM.2017.8072767.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 August 2023 doi:10.20944/preprints202308.0963.v1

17. Z. Jiang, Z. Wang, L. Nie, X. Zhao, and S. Huang, "A Low-Profile Ultrawideband Slotted Dipole Antenna Based on Artificial Magnetic Conductor," *Antennas Wirel. Propag. Lett.*, vol. 21, no. 4, pp. 671–675, Apr. 2022, doi: 10.1109/LAWP.2022.3140802.
18. M. M. Hossain, M. J. Alam, and S. I. Latif, "Orthogonal Printed Microstrip Antenna Arrays for 5G Millimeter-Wave Applications," *Micromachines*, vol. 13, no. 1, p. 53, Dec. 2021, doi: 10.3390/mi13010053.
19. P. Kumar *et al.*, "An Ultra-Compact 28 GHz Arc-Shaped Millimeter-Wave Antenna for 5G Application," *Micromachines*, vol. 14, no. 1, p. 5, Dec. 2022, doi: 10.3390/mi14010005.
20. M. N. Hasan, S. Bashir, and S. Chu, "Dual band omnidirectional millimeter wave antenna for 5G communications," *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 12, pp. 1581–1590, Aug. 2019, doi: 10.1080/09205071.2019.1617790.
21. A. N. S. Shamsuri Agus *et al.*, "Combined RIS and EBG Surfaces Inspired Meta-Wearable Textile MIMO Antenna Using Viscose-Wool Felt," *Polymers*, vol. 14, no. 10, p. 1989, May 2022, doi: 10.3390/polym14101989.
22. C. A. Balanis, *Antenna theory: analysis and design*, 3rd ed. Hoboken, NJ: John Wiley, 2005.
23. X. Shen, Y. Liu, L. Zhao, G.-L. Huang, X. Shi, and Q. Huang, "A Miniaturized Microstrip Antenna Array at 5G Millimeter-Wave Band," *Antennas Wirel. Propag. Lett.*, vol. 18, no. 8, pp. 1671–1675, Aug. 2019, doi: 10.1109/LAWP.2019.2927460.
24. K. Bangash, M. M. Ali, H. Maab, and H. Ahmed, "Design of a Millimeter Wave Microstrip Patch Antenna and Its Array for 5G Applications," in *2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, Swat, Pakistan: IEEE, Jul. 2019, pp. 1–6. doi: 10.1109/ICECCE47252.2019.8940807.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Journal of the Indian Society of Remote Sensing
https://doi.org/10.1007/s12524-023-01734-9

RESEARCH ARTICLE

A Cross-Domain Semi-Supervised Zero-Shot Learning Model for the Classification of Hyperspectral Images

Pallavi Ranjan[1] · Gautam Gupta[2]

Received: 17 February 2023 / Accepted: 14 June 2023
© Indian Society of Remote Sensing 2023

Abstract
The cost of obtaining spectral information has dropped significantly due to the remarkable progress in hyperspectral imagery. However, annotating the hyperspectral samples remains a labour-intensive and time-consuming behaviour. Given the challenge of labelling annotations, most conventional hyperspectral classifiers are tested and trained to use a solitary hyperspectral image cube, resulting in two challenges. Firstly, the existing classification models give reasonably close to perfect accuracy, but they are hard to extrapolate to other datasets. Second, hyperspectral datasets are not typically collected in the same scene; different data sets will contain distinct spectral bands. To address these concerns, a new framework for hyperspectral image stratification is proposed in this paper that requires training and analysing separately throughout heterogeneous hyperspectral datasets. Even though annotating hyperspectral data are beneficial, it has received very little attention from the hyperspectral community. The proposed zero-shot learning-based novel model maps the features between training and testing datasets by proposing a convolutional and word embedding model. The feature mapping module associates the training and testing set leading to classification. Experiments on benchmark hyperspectral datasets including Indian Pines (IP), Pavia University (PU), Kennedy Space Center (KSC) validate the performance of proposed cross-domain model.

Keywords Hyperspectral classification · Zero-shot learning · Cross-domain · Deep learning

Introduction

A hyperspectral image (HSI) is a three-dimensional cuboid with an improved spectral resolution than traditional remote sensing imagery. Amongst other areas, HSI classification is one of the prominent domains in hyperspectral data observation. It refers to classifying each pixel of an image based on spectral and spatial properties. The rich spatial and spectral data inside an HSI improve the interpretation potentiality of remote sensing imagery, attracting interest and use in environmental monitoring, medical, defence, military, and precision agriculture, to name a few (Li et al., 2019; Lv & Wang, 2020).

Due to the rich data in HSI, substantial work has been developed to leverage existing spatial and spectral information (Paoletti et al., 2018) effectively. Because trying to label hyperspectral data is time-consuming and expensive, the current hyperspectral realm lacks a sized directory such as ImageNet (Deng et al., 2009). Furthermore, a small percentage of datasets have labelled pixels, whereas conventional supervised models rely on extensive labelled training data for good classification performance (Wu & Prasad, 2018). This challenge of insufficient labelled samples jeopardises the development of supervised and semi-supervised algorithms (Pan et al., 2021). Moreover, datasets like Indian Pines (IP) and Salinas (SA) (Pan et al., 2021) characterize themselves in terms of spatial distribution and land cover. Trained models generalise poorly to unseen datasets resulting in a poor classification performance (Lampert et al., 2009; Pan et al., 2021; Wu & Prasad, 2018).

To address the above-mentioned limitations, we present a new framework for HSI classification that can perform

✉ Pallavi Ranjan
er.pallaviranjan@gmail.com

[1] Delhi Technological University, New Delhi, India

[2] Indian Institute of Technology Madras, Chennai, India

Published online: 31 July 2023

cross-domain classification. The training set for the proposed novel model will be disjoint from the testing set. Evaluation of an entirely new dataset would result in unseen and unknown categories of data. Hence, the critical challenge is establishing associations between seen and unknown data, that is, disjoint training and testing sets. Hyperspectral data labels are arbitrarily defined using natural science-known facts. Label semantic representations can be viewed as a type of high-level feature. Inspired by recent advances in zero-shot learning (Biswas & Annadani, 2018; Lampert et al., 2014; Pan et al., 2021; Socher et al., 2013), we create word embeddings to bridge the gap between seen and unseen categories. The feature mapping in the embedding space is used to obtain the relationships between trained categories in the training set and unseen categories in the testing set. As a result, the salient points of this study are:

1. This study proposes a cross-domain classifier, which differs from the homogeneous idea of training and testing on the same datasets. The newly proposed novel HSI classifier categorises datasets captured using different sensors.
2. The proposed classification framework builds a novel model trained using similarity mapping between extracted features and generated word embeddings. The features are extracted using a convolutional neural network (CNN), and word embeddings are generated using long short-term memory (LSTM)-based module. The pre-trained model is tested on a dataset with unseen classes.
3. The trained model was tested against three benchmark hyperspectral datasets to validate its performance.

Related Work

Table 1 summarises a list of related cross-domain models based on zero-shot learning and summarises their performance.

Methodology

The proposed model consists of two phases. First, a pre-trained feature mapping module $f_m(.)$ is built using the visual features $f(V)$ and word embeddings $g(E)$ present in the training dataset. In the second phase, the testing dataset is passed through a pre-trained feature-mapped module $f_m(.)$ to assess the similarity , between training data τ_r and testing τ_ε data and predict a label b_j^{te}. All experiments were run using 12 GB RAM on the TensorFlow framework, and GPU was 1×Tesla K80 compute 3.7 2496 CUDA cores 12 GB GDDR5 VRAM. The optimized learning rate was 0.00001, and the AdaDelta optimizer ran for 60 epochs.

Feature Mapping Model

The training HSI data are embedded into hyperspectral visual features using an HSI classification model (Firat et al., 2022). The significant spectral and spatial features contained in HSI cubes were retrieved using the three dimensional CNN (3D-CNN) and ResNet model to tackle the problem of vanishing gradients. Mei et al. (2019) combined an attention network with CNN to focus on the discriminative features, eventually dealing with the problem of the complexity of HSI data. Multiple other models are available that deal with HSI classification (Dang et al., 2023; Ranjan & Girdhar, 2022). CNN has excellent potential in automatically learning the features. However, it only captures the spatial features and disregards the essential spectral features contained in HSI (Guo et al., 2022). Moreover, the accuracy decreases with an increase in the number of layers (Zhong et al., 2017). Therefore, to extract the visual features, this study uses the residual network introduced by Socher et al. (2013), which has shown excellent performance for classifying HSI by mitigating the declining accuracy phenomenon (Zhong et al., 2017). The main feature of ResNet is layer skipping, which effectively alleviates the gradient descent problem by reusing previous layers activations until the subsequent layer learns its weights (Alotaibi & Alotaibi, 2020; Xue et al., 2021). The residual network can be described by Eq. (1) where A_i is the output of ith unit, W_i denotes the parameters of residual structure. The nonlinear layers as shown in Fig. 1, the nonlinear layers stacked together build the function $F(A_i, W_i)$ instead of mapping desired A_{i+1} directly. Deep residual networks are easier to optimize then CNNs, have more representative capacities and provide higher classification accuracy considering the deeper layers involved (Zhong et al., 2017). This model was optimized using adaptive moment optimization and sparse categorical cross entropy, trained for 60 epochs and batch size 16 with verbose set to true. Figure 1 depicts the process.

$$A_{i+1} = \max\{0, A_i + F(A_i, W_i)\} \quad \text{(Zhong etal. 2017)} \qquad (1)$$

Word Embedding Model

The word embeddings were generated to connect the training and testing set that are disjoint sets. Various methods exist in the literature that establishes relationships

Table 1 Current trends in literature including proposed models and their performance

References	Model	Hardware	Datasets used and OA				Findings	Limitations
			DA	OA	AA	KC		
Zhong et al. (2018)	SSRN	MSI GT72S with GTX 980 M GPU	IP	61.36	59.75	56.91	Alleviates the declining accuracy phenomenon by adding residual connection; learns spectral and spatial features consecutively; batch normalization is used as a regularization method	It gives excellent performance when trained and tested on the same dataset. When cross-domain was applied (Li et al., 2022), the accuracy was low
			PU	76.26	79.51	70.56		
			SA	86.39	93.24	84.95		
Wang et al. (2023)	GMTN	TensorFlow framework	PU	84.97	80.73	79.77	The proposed study alleviates the problem of limited labelled samples using domain adversarial training	The study is semi-supervised that resolves the limited samples problem. However, performance can be improvised
			IP	68.36	93.62	64.03		
			SA	91.80	74.04	90.88		
Zhang et al. (2022)	DCFSL	PyTorch, Intel Core i7-8700 with NVIDIA GTX 1080ti GPU	Tr/Te	Chikusei	Houston		They have proposed a model that works using few-shot learning and cross-domain adaptation. This model classifies target classes with a few labelled samples. The features extracted are domain independent and overcome the domain shift problem	The proposed model trained on Chikusei and Houston datasets provided great classification accuracy on Pavia Centre and SA; however, the accuracy for PU and IP dataset can be increased
			PU	83.65	82.38			
			PC	97.45	96.86			
			SA	89.34	89.14			
			IP	66.81	60.88			
Gao et al. (2020)	RN-FSC	Pytorch, Intel Core i7-9750H, with Nvidia GeForce RTX 2070	PU	$L = 5$	81.94	61.74	This study proposes a cross-domain model that classifies the target dataset with just a few labelled samples. The proposed model incorporates a feature learning module and a related learning module that identifies spectral-spatial information	The proposed method outperforms traditional machine learning methods like semi-supervised SVM and other semi-supervised deep learning methods. There is a scope to discover more ways to find discriminative spectral-spatial features and relationships between them to improve classification performance
			PU	$L = 25$	91.05	80.22		
			PC	$L = 5$	96.36	85.63		
			PC	$L = 25$	98.72	92.10		
			SA	$L = 5$	86.99	71.08		
			SA	$L = 25$	92.15	80.45		
Zhu et al. (2021)	DSAN	GeforceGTX 1080Ti GPU	IP	60.74		56.69	This study uses subdomain adaptation that explicitly captures the dependencies between categories with the same attributes instead of focusing on global dependencies	The convergence of the proposed network was too slow because of several loss functions contained in it. When performed on hyperspectral datasets by Li et al., (2019), the performance classification was low
			PU	78.43		70.58		
			SA	85.35		83.66		
Yu et al. (2019)	DAAN	PyTorch framework	PU	76.54		68.10	This study proposed a domain adaptation network having a source and target network as a disjoint set. This network dynamically learns the variance between two sets	This network achieved low classification accuracy when applied to a complex problem set by Li et al., (2019), like hyperspectral. The accuracy for SA was achieved higher than PU or IP
			IP	57.45		52.66		
			SA	84.06		82.23		
Zhang et al. (2022)	Gia-CFSL	PyTorch, Intel Core i7-8700 with NVIDIA GTX 1080ti GPU	PU	84.47		79.71	This study proposed a novel model combining domain adaptation with few-shot learning and calculates intradomain similarity and cross-domain similarity	Trained network using Chikusei and PU and tested using IP, PU and SA. This network provided low accuracy for Indian Pines
			IP	67.42		63.10		
			SA	89.95		88.82		

Table 1 continued

References	Model	Hardware	Datasets used and OA				Findings	Limitations
			DA	OA	AA	KC		
Liu et al., 2021b)	MDL4OW/C	TensorFlow 1.9, Keras 2.1.6, Python 3.6, Intel I5-8500 CPU, an Nvidia GTX 1080 Ti GPU	PU	L = 20	85.07	85.05	This study proposed a method to classify novel and unknown classes in the unseen set. The unknown and known classes are differentiated using a threshold. This study differed from others because it was based on the assumption that unseen classes could also belong to a novel set	Performance Classification was excellent for close sets; however, more experiments can be conducted to check the open set and if classification performance can be increased
			PU	L = 200	90.07	91.64		
			SA	L = 20	87.40	82.96		
			SA	L = 200	93.73	95.37		
			IP	L = 20	68.92	71.68		
			IP	L = 200	82.23	83.74		
Xi et al. (2022)	CMFSL	NVIDIA GeForce GTX 1080Ti GPU	IP	67.40	80.05	63.29	This few-shot learning-based method performs classification using a class covariance matrix. The proposed novel feature extractor reduces domain shift between the disjoint training and testing set	While computing similar classes, Mahalanobis distance provided better results than Euclidean or Cosine distance
			SA	90.26	94.93	89.18		
			PU	82.75	85.90	77.96		
			PC	97.21	94.17	96.06		
Liu et al., (2021a)	DMVL	Pytorch, Intel i7-9750, Nvidia GeForce RTX 2070 M	PU	85.70	81.22	81.07	This study constructs latent space using ResNet trained with augmented views of the same dataset. This network extracts view-invariant features to achieve excellent classification	The proposed network achieves great classification accuracy. However, it is computationally expensive due to the time taken to train the network
			IP	73.95	84.82	71.07		
			SA	95.88	96.49	95.43		
			HU	76.65	78.95	75.45		

Dataset (DA), Overall accuracy (OA), Average accuracy (AA), Kappa Coefficient (KC) for Indian Pines (IP), Pavia University (PU), Salinas (SA), Pavia Center (PC), Houston (HU) showing Training dataset (Tr) and Testing dataset (Te) proportion

between categories using semantic information like attributes (Biswas & Annadani, 2018). Other methods that use high-level attributes to discover the semantic relationships are presented by Lampert et al., (2009, 2014). A study by Akata et al. (2016) embeds classes using the attribute vector space. The parameters of this component are found on a training dataset of labelled samples to ensure that the proper classes score higher than the erroneous ones in a given image. Wang et al. (2016) mapped the image and text description into a common vector space and computed similarity using cosine distance. Unlike CNN models, the LSTM focuses on assessing the temporal structure beginning with words or characters (Reed et al., 2016). This is an LSTM-based model that generates text vectors. Given a text description of w words $T = \{t_1, t_2, t_3 \ldots t_w\}$, the proposed LSTM-based model models the long-term dependencies using two types of states, including a cell state defined as c and a hidden state defined as h. The existing inputs x_t and previous state inputs (c_{t-1}, h_{t-1}) are combined for a specific time step to compute the current state. Three gates control the message passing during the integration, including a forget gate f_t, an input gate i_t, and an output gate o_t.

The proposed Conv1D-LSTM embedding model works using vector-based similarity representation, allowing greater capacity to characterise global alignments between images and sentences. The text data containing the description of classes, as shown in Table 1 were collected from Graña et al. (2021). The collected data were pre-processed using lemmatization, removing stop words, contraction mapping, and tokenization technique, and the word embeddings were generated and fed into the embedding layer. The input is a textual description of the label, and the output is a vector describing the sequence. This vector is then compared to a new label vector to determine which label is most likely to be similar to the image. The proposed embedding Conv1D-LSTM model, depicted in Fig. 2, employs an embedding layer followed by Conv1D and max-pooling, which is then fed into an LSTM and flattening layer to classify the word embeddings and produces vector output. This was optimized using adaptive moment optimization and sparse categorical

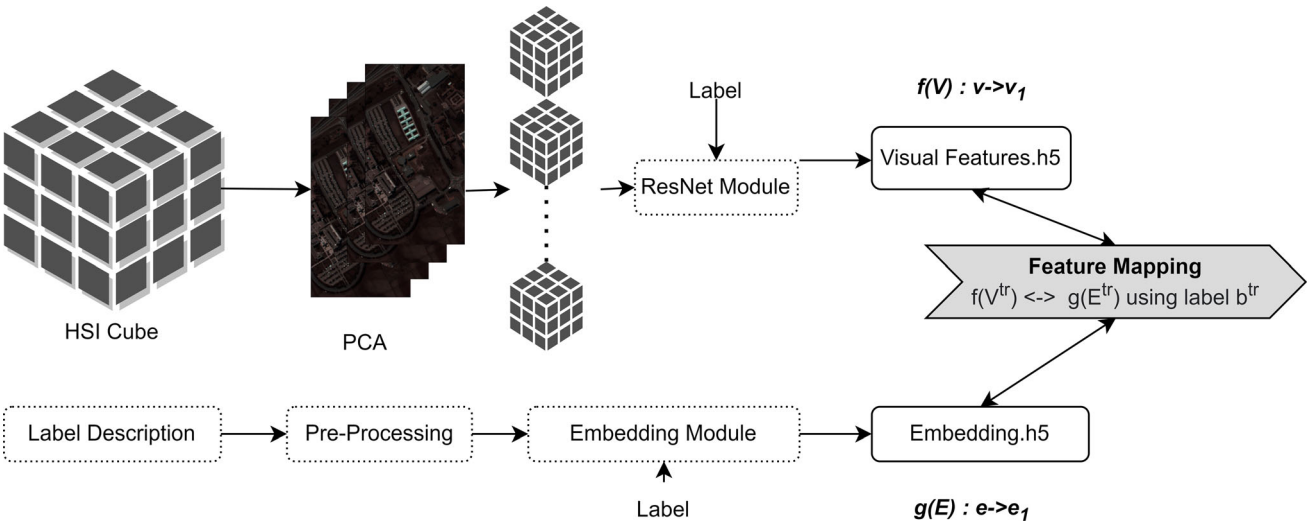


Fig. 1 Building a feature mapping model in Python 3.9 using TensorFlow framework

entropy, trained for 60 epochs in a batch size of 16 with verbose set to true. Figure 1 depicts the process briefly, and Fig. 2 depicts the layers involved in the embedding model.

Layout Similarity Mapping

The way features are mapped with semantic embeddings differs in different zero-shot learning-based networks. Various models discover projection mapping from visual to semantic space (Lampert et al., 2014; Socher et al., 2013), whereas others learn projections from semantic to visual space (Shigeto et al., 2015). This study (Pan et al., 2021) proposed a method for hyperspectral data that maps label embeddings to the visual feature space and learns correspondence between semantic visual embeddings in the visual space. It is done to ameliorate the hubness issue typically found in nearest neighbour searches for high-dimensional information (Radovanovic et al. n.d.). The proposed work maps visual features with word embeddings using a standard label y^{tr}. Figure 3 shows how the similarity is mapped. Table 2 shows the description of labels.

Testing the Pre-trained Model

The proposed work maps visual features with word embeddings using a standard label y^{tr}. The testing data were reduced to hundred bands by selecting discriminative bands using principal component analysis and eventually fed into a pre-trained feature mapping module. The pre-

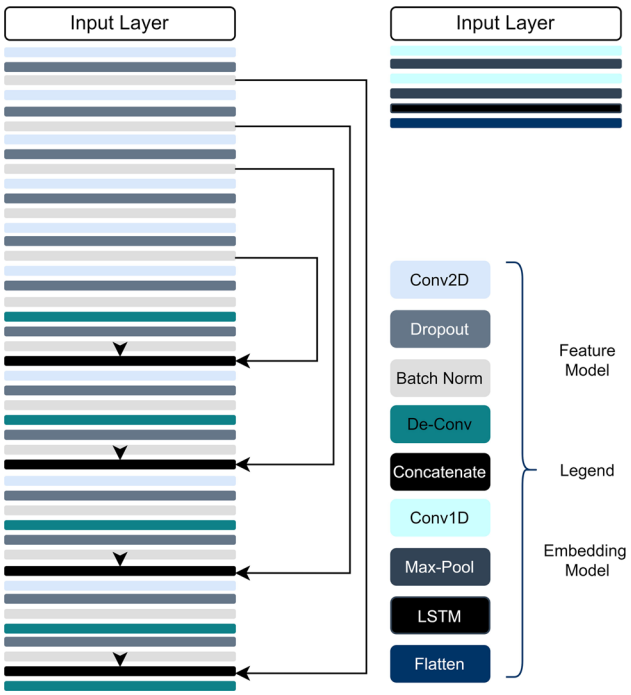


Fig. 2 Layers involved in feature model and embedding model

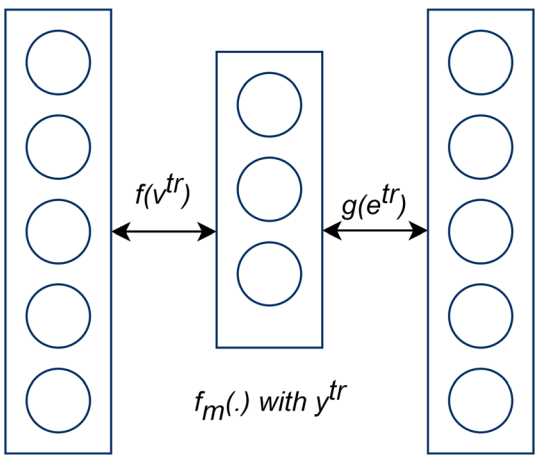


Fig. 3 Mapping similarity

Table 2 Descriptions of categories

Asphalt	Sticky; black; extremely viscous liquid petroleum in semi-solid form; used as a binder in road construction; used for sealing flat roofs
Meadows	An open habitat; is a field with grasses, herbs, and other non-woody plants for vegetation. Demonstrates animal courtship insects that nest, gather food, and pollinate; meadows that are agricultural, transitional, or perennial
Gravel	Fragments of rock; crushed stone; fine, medium, or coarse; used for road construction, construction fill, and a variety of other minor purposes; distributed as river deposits, river flood plains, and glacial deposits
Painted metal sheets	A type of metal; used in metalworking; thin sheets used as foil or leaf, and thick pieces used as a plate or structural steel; other metals like aluminium, copper, nickel, silver, and gold can be used as metal sheets too

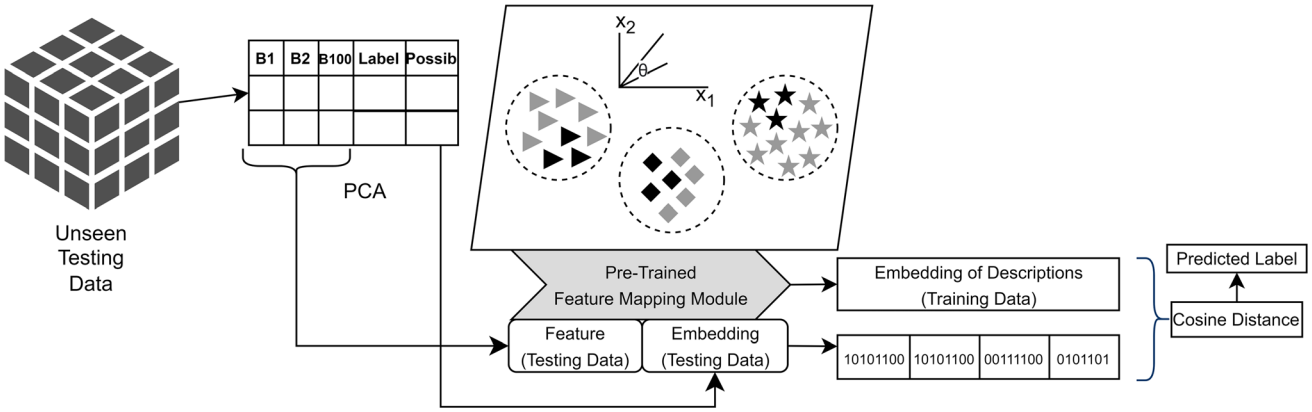


Fig. 4 Testing the model

trained feature mapping is transmitted to the visual elements of test samples and word embeddings of the testing categories. A metric learning method assigns the most appropriate category. The proposed model uses cosine similarity computed using a similarity vector between different training and testing datasets. The cosine or Euclidean distance is used in most previous methods to represent the similarities of two feature vectors that can record the relevance of the testing point with all the training points. The metric was top-K classification accuracy, and the results were top-4 accuracy. Before dividing the total number of classes, the average of correct predictions was calculated for each class. The accuracy of multiclass classification was calculated by evaluating the given test image on text descriptions from all classes, sorting the final prediction score in increasing order and selecting the top accuracy, algorithm 2 for reference. For $k = 4$, the top-k accuracy is written in Eq. (2), where m_c is the number of correct predictions of top-k rank in c. Figure 4 describes the process.

$$\text{accuracy}_{\text{topk}} = \frac{1}{||\tau_\varepsilon||}\sum_{i=1}^{||\tau_\varepsilon||}\frac{m_c}{\left\{b_j^{\text{te}}\right\}} \tag{2}$$

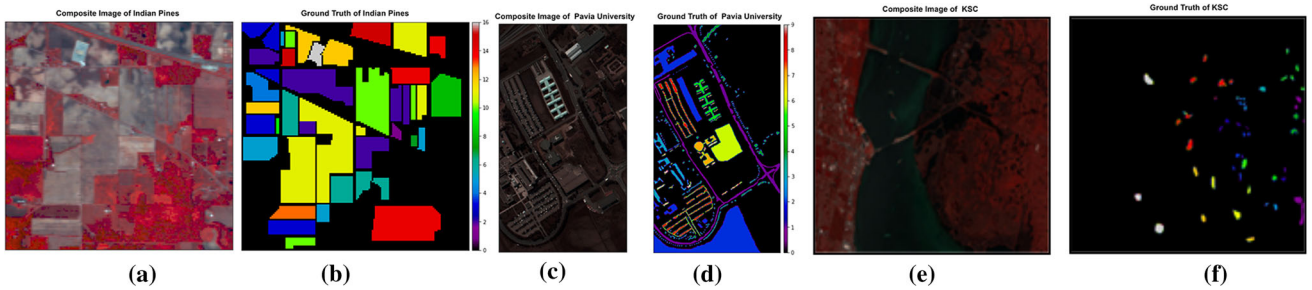


Fig. 5 **a** Composite image of Indian Pines **b** ground truth of Indian Pines **c** composite image of Pavia University **d** ground truth of Pavia University **e** composite image of Kenedy space centre **f** ground truth of Kennedy space centre

Algorithm 1 : The proposed model

Input :

Seen Training Images $\tau_r = \{a_i^{tr}, b_i^{tr}\}$ for *i=1 to n* in τ_r

Unseen Testing Images $\tau_\varepsilon = \{a_j^{te}\}$ for *j=1 to m* in τ_ε

$\mathbb{Z}\tau_r$ and $\mathbb{Z}\tau_\varepsilon$ Classes belonging to seen and unseen data

Output :

Predicting labels $\{b_j^{te}\}$ from unseen testing data for *j=1 to m* in τ_ε

Step 1. Finding visual features using Feature Modelling $f(V): v \Longrightarrow v_1$

Step 2. Find word embeddings using Embedding Modelling $g(E): e \Longrightarrow e_1$

Step 3. **for** $\forall \{a_i^{tr}, b_i^{tr}\}$, find feature mapping $f_m(.): f(V^{\tau_r}) \Leftrightarrow g(E^{\tau_r})$

$f_m(.): v_1 \longrightarrow e_1$ on τ_r with label b_i^{tr}

$f_m(.): e_1 \longrightarrow v_1$ on τ_r with label b_i^{tr}

end

Step 4. Similarity Prediction

for $\forall a_j^{te} \in \tau_\varepsilon$ Repeat below steps

Get the $f(a_j^{te}) \in v_1$ and $g(e_j^{te}) \in e_1$

Using the pre-trained $f_m(.)$, get the embedding

Using cosine similarity $cos(.)$, Find the final label b_j^{te}

end

Algorithm 2 : Sorting the prediction score

```
for i=1 to n-1
        j=1
        do while j>0 and A[j]<A[j-1]
        T = A[j]
        A[j] = A[j-1]
        A[j-1] = T
        j = j-1
        end
end
```

Experimental Settings

In order to test the performance of the proposed model, three datasets were chosen, including Pavia University (PU), Indian Pines (IP), and Kennedy Space Centre (KSC) (Graña et al., 2021). The composite and ground truth image of these datasets are shown in Fig. 5. Table 3 shows the properties of these datasets, including the sensor using which these were captured. The bands collected by the sensor is specific to the sensor type due to difference in wavelength and spectral resolution. The experimental dataset was consistent by selecting hundred discriminative bands from the dataset. We have taken three datasets for experimental purposes and divided them into two groups. For group 1, we trained the model using PU and IP and KSC, whereas Group 2 was trained using IP and tested using PU and KSC. In T_r T_e, T_r denotes the dataset on which the model was trained, and T_e denotes the testing set on which results were obtained. Figure 6 shows the sample distribution for each dataset.

Datasets

Results

Figure 7 shows the spectra for the IP dataset. This presents the interclass and intraclass distance between different categories of IP for particular bands. Figure 8 shows the PCA projection of IP classes over PU classes. The coloured dots represent the PU classes, and the black dots represent the IP classes. Some classes are hard to classify due to similar projections. For instance, there is an overlap in Grass-trees and Grass-pasture. Also, other classes like Soybean-clean and Soybean-notill are had similar PCA 2D projections. Additionally, there was overlap found in Corn-mintill, Corn-notill, and Alfalfa. This overlap is due to the similar spectra of these classes that can cause misclassification. The added word embeddings reduce this misclassification by having a good separability of classes.

Table 3 Dataset description

	Pavia University (PU)	Indian Pines (IP)	Kennedy Space Center (KSC)
Spatial size	610 × 340 pixels	145 × 145 pixels	512 × 614 pixels
Spectral range	430 to 860 nm	400–2500 nm	400–2500 nm
Bands	103	224	176
Sensor	ROSIS	AVIRIS	AVIRIS
Areas covered	Pavia	Indiana	Florida
Categories	9	16	13

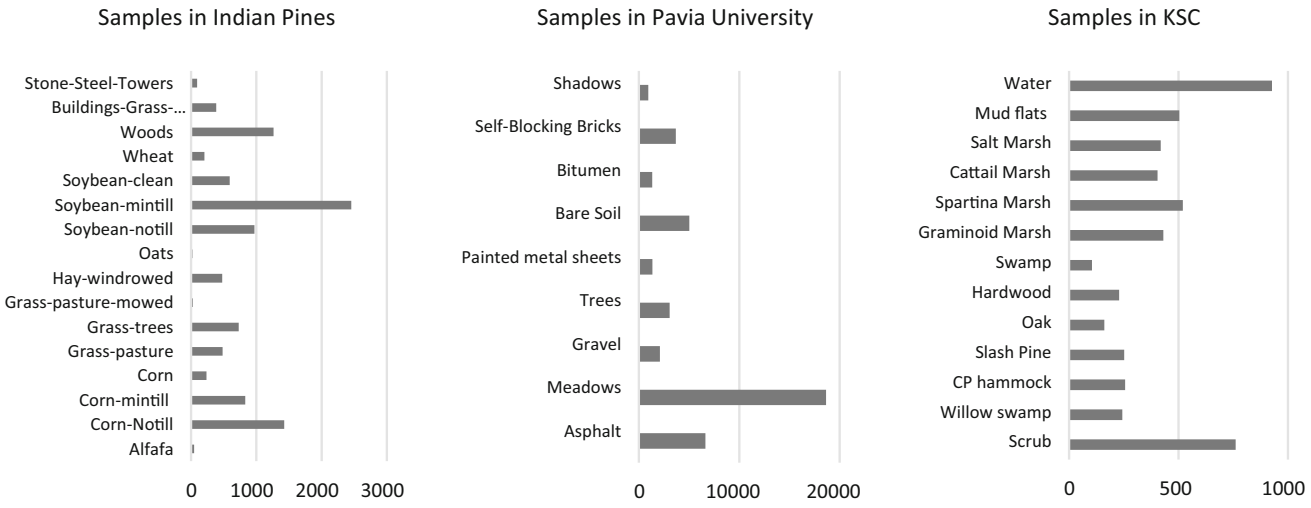
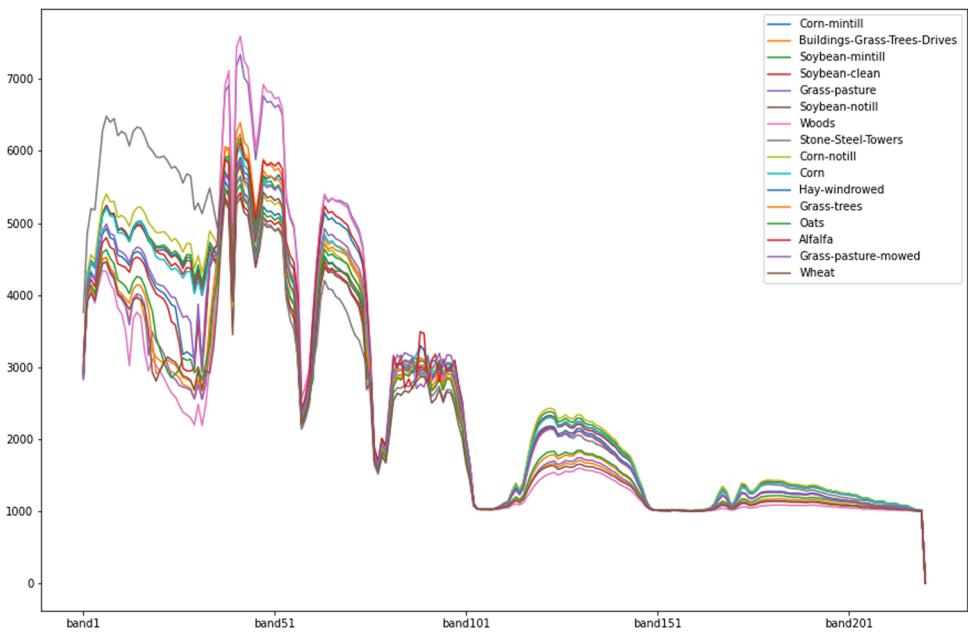


Fig. 6 Division of samples for each dataset

Fig. 7 Indian Pines spectra



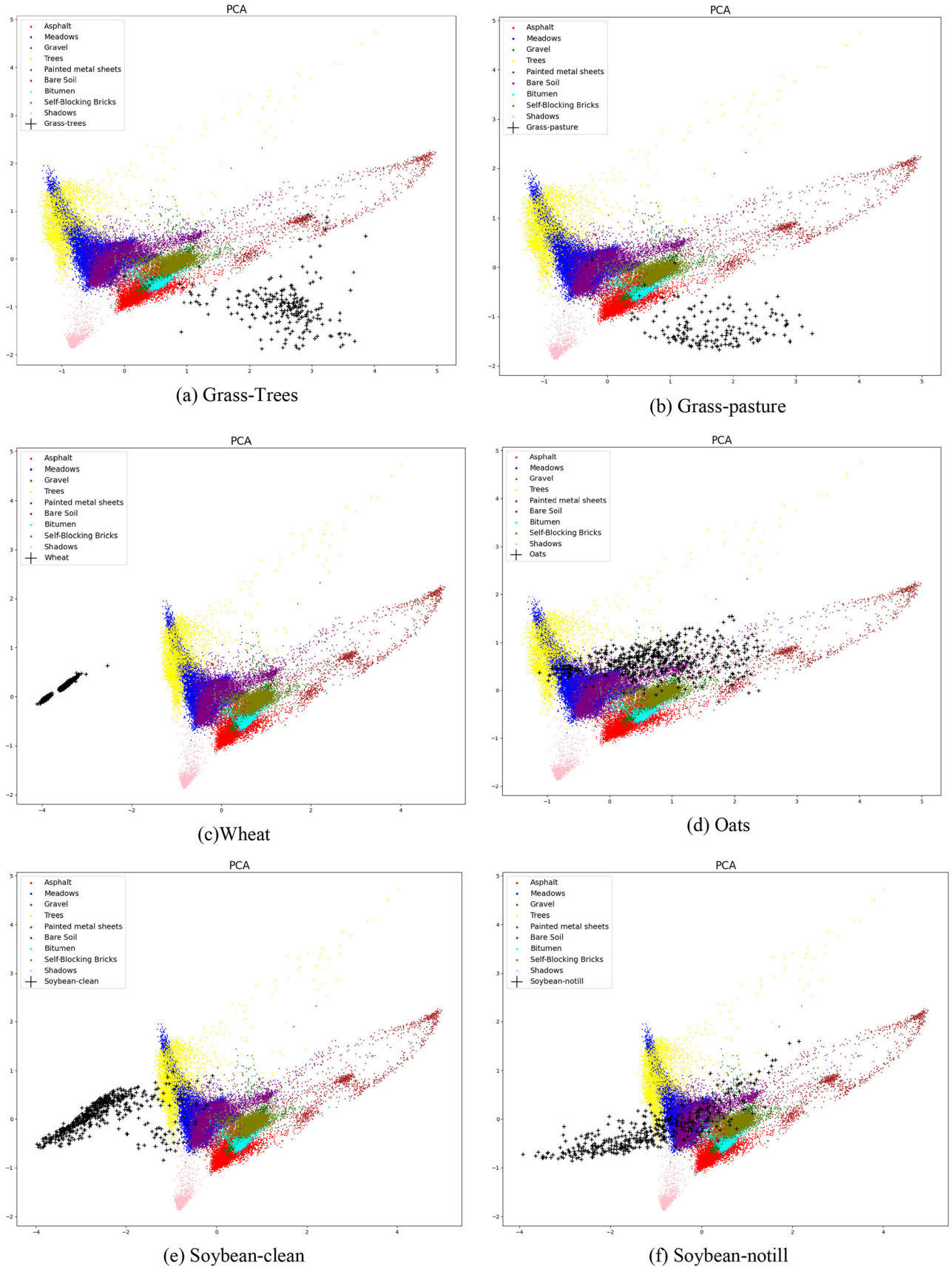
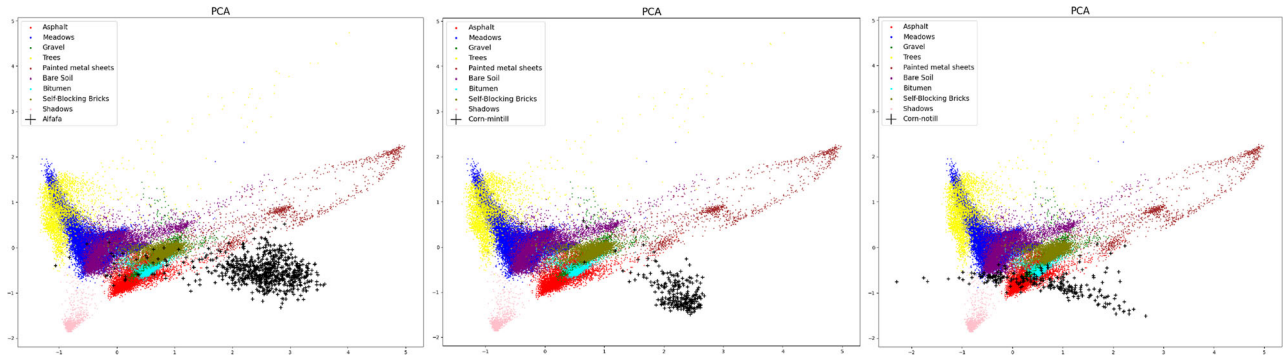


Fig. 8 Principal component analysis (PCA) projection of Indian Pines over Pavia University



(g) Alfafa (h) Corn-mintill (i) Corn-notill

Fig. 8 continued

Table 4 Classification accuracies achieved with T_r (PU) T_e (IP) with best values in bold

	SSRN (Zhong et al., 2018) in Li et al., (2022)	RN-FSC (Gao et al., 2020) in Li et al., (2022)	DSAN (Zhu et al., 2021) in Zhang et al., (2022)	DAAN (Yu et al., 2019) in Zhang et al., (2022)	DFSL + NN (Li et al., 2022)	DFSL + SVM (Li et al., 2022)	DCFSL (Zhang et al., 2022)	Gia-CFSL (Zhang et al., 2022)	Proposed method
1	18.38	96.34	**97.83**	39.13	96.75	96.75	95.37	91.60	50.00
2	64.79	46.13	28.29	26.82	38.65	36.38	43.26	50.76	**80.14**
3	27.65	40.61	**94.70**	90.96	42.79	38.34	57.95	44.88	52.61
4	26.97	58.62	**97.89**	91.14	68.10	77.16	80.60	81.61	67.60
5	80.76	64.96	**90.89**	68.53	71.20	73.92	72.91	70.76	84.82
6	86.87	69.45	38.08	13.29	76.18	86.25	87.96	84.25	**99.54**
7	32.24	100	96.43	25.00	**100**	97.10	99.57	97.10	87.50
8	100	77.70	81.59	59.83	74.84	81.82	86.26	91.12	**100.00**
9	57.69	**100**	**100**	20.00	**100**	75.56	99.33	99.26	83.33
10	59.69	25.49	52.37	49.49	47.98	52.22	62.44	**62.68**	38.01
11	70.87	65.51	38.17	45.87	57.95	59.96	62.75	66.54	**83.31**
12	45.00	27.13	**96.12**	91.57	38.21	36.56	48.72	42.06	62.92
13	88.29	99.75	14.63	15.12	97.50	98.00	**99.35**	97.11	96.72
14	97.18	76.35	98.66	**99.60**	83.44	84.63	85.40	87.10	93.94
15	36.64	70.34	57.51	67.36	62.29	**74.10**	66.69	68.74	69.82
16	60.98	**100**	94.62	96.77	**100**	**100**	97.61	97.47	96.42
OA	61.36	58.17	60.74	57.45	59.65	61.69	66.81	67.42	**96.23**
AA	59.75	69.90	73.61	56.28	72.24	73.05	77.89	77.06	**77.92**
KC	56.91	52.52	56.69	52.66	54.55	56.78	62.64	63.10	**94.25**
Training time									NA
Testing time									25

Category 1: T_r (PU) → T_e (IP)

The first category included training on PU and testing on IP using the proposed model. As shown in Table 4, The models that were used for comparison included SSRN (Li et al., 2022), RN-FSC (Li et al., 2022), DSAN (Zhang et al., 2022), DAAN (Zhang et al., 2022), DFSL + NN (Li et al., 2022), DFSL + SVM (Li et al., 2022), DCFSL (Zhang et al., 2022), and Gia-CFSL (Zhang et al., 2022). All these models were trained and tested using cross-domain data with testing results on IP included in Table 4. Bold values depict the best results for each category. The

Table 5 Classification accuracies achieved with T_r (IP) T_e (PU) with best values in bold

	RN-FSC (Gao et al., 2020)	DSAN (Zhu et al., 2021) in (Zhang et al., 2022)	DAAN (Yu et al., 2019) in (Zhang et al., 2022)	MDL4OW/C (Liu et al., 2021b)	CMFSL (Xi et al., 2022)	DCFSL (Zhang et al., 2022)	Gia-CFSL (Zhang et al., 2022)	Proposed method
1	87.28	91.15	**93.38**	94.54	76.35	81.06	81.16	79.30
2	84.33	94.10	95.10	**98.81**	82.28	87.74	89.58	05.68
3	90.42	29.16	43.26	**96.72**	71.63	63.33	59.77	56.32
4	78.09	96.90	92.30	90.61	93.54	92.56	93.17	**98.52**
5	99.56	99.63	99.63	97.47	**99.86**	99.01	99.46	89.46
6	63.25	27.36	54.32	**86.29**	85.00	74.58	76.95	56.33
7	52.09	58.35	11.28	**80.56**	85.59	77.74	79.58	60.10
8	84.81	54.02	0.46	**95.74**	79.38	62.43	74.36	80.66
9	95.94	94.72	88.38	94.52	**99.46**	98.22	98.56	64.38
OA	81.94	78.43	76.54	90.53	82.75	82.40	84.47	**90.42**
AA	81.75				**85.90**			65.64
KC	75.84	70.48	70.58		77.96	77.11	79.71	**89.62**
Training time								NA
Testing time								45

Table 6 Classification accuracies achieved with T_r (IP,PU) T_e (KSC) with best values in bold

KSC						T_r (PU), T_e ((KSC)	T_r (IP), T_e (KSC)
	RN-FSC (Gao et al., 2020) in (Chi et al., 2022)	DFSL (Liu et al., 2019) in (Chi et al., 2022)	SVM (Chi et al., 2022)	DMVL (Liu et al., 2021a) in (Guan & Lam, 2022)	3DCNN (Chi et al., 2022)	Proposed method	Proposed method
1	54.07	98.45	58.32	92.62	48.18	**99.60**	72.75
2	72.77	90.60	41.22	72.43	72.69	**94.23**	51.67
3	58.25	82.92	48.25	83.98	55.58	**88.67**	38.83
4	49.07	62.17	18.51	14.29	31.68	**68.65**	44.53
5	68.57	46.40	27.23	**96.89**	33.81	72.67	87.55
6	58.30	**68.85**	21.05	27.95	15.18	58.08	37.89
7	**100.00**	61.40	43.03	**100.00**	80.50	90.47	74.31
8	87.19	76.84	35.36	77.99	54.40	**98.14**	34.35
9	60.24	92.64	66.06	77.89	89.03	**97.88**	29.95
10	84.02	97.71	33.53	94.80	53.07	**98.51**	71.44
11	54.33	**100.00**	83.26	90.22	93.78	98.32	24.97
12	93.03	96.23	29.32	90.06	49.30	**98.01**	27.82
13	99.25	99.28	83.96	100.00	87.74	**100.00**	69.22
OA	70.62	87.77	56.13	83.20	63.49	**92.4**	83.76
AA	72.24	82.58	48.88	78.39	58.84	89.48	51.17
KC	67.02	86.42	51.13	81.35	59.79	89.43	81.68
Training time						NA	NA
Testing time						18	65

Table 7 Classification accuracies achieved with T_r (PU) T_e (PU), T_r (IP) T_e (IP) with best values in bold

	T_r (PU), T_e (PU) Proposed method	T_r (IP), T_e (IP) Proposed method
1	98.61	93.49
2	99.01	80.25
3	96.35	98.94
4	98.21	99.63
5	99.92	98.65
6	99.14	72.33
7	90.85	93.29
8	94.54	98.58
9	99.89	98.58
10		93.46
11		88.52
12		86.97
13		99.91
14		97.55
15		99.45
16		83.01
OA	**99.49**	**93.46**
AA	**97.39**	**92.66**
KC	**98.26**	**90.43**
Training time	36,000	74,500
Testing time	12	72

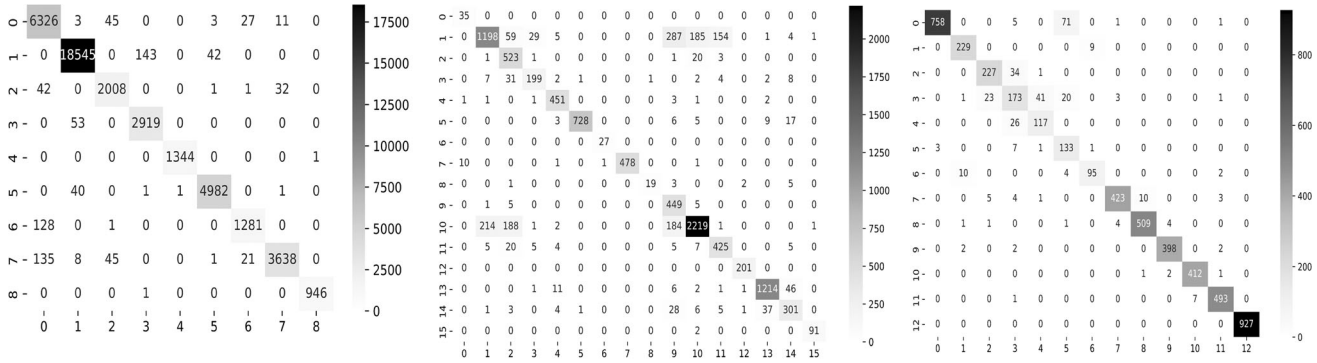
Table 8 Number of parameters involved while training convolutional and embedding model

Conv. model	Embedding model
Total params: 7,959,809	Total params: 61,824
Trainable params: 7,955,073	Trainable params: 61,824
Non-trainable params: 4736	Non-trainable params: 0

comparison was made using overall accuracy, average accuracy and kappa coefficient as the performance metrics. The proposed model achieved the best results in Corn-notill, Grass–Trees, Hay-windrowed and Soybean-mintill, having comparatively a higher number of samples. The classes achieving very low overall accuracy were Alfafa, Corn-mintill and Soybean-Notill. The highest overall accuracy, average accuracy and kappa coefficient was achieved by the proposed model validating the performance of the proposed model. The mapping of feature and word embeddings had a considerable effect on the performance of the model. Despite the fact that IP and PU have different category distribution, the spectral-spatial model was able to attain excellent classification accuracy. However, mining the HSI information carefully is an important task.

Category 2: T_r (IP) → T_e (PU)

The second set of experiments was done by training the proposed model with IP and testing with PU. The models that were used for comparison were cross-domain tested using the PU dataset, including RN-FSC (Gao et al., 2020), DSAN (Zhu et al., 2021) in Zhang et al. (2022), DAAN (Yu et al., 2019) in Zhang et al. (2022), MDL4OW/C (Liu et al., 2021b), CMFSL (Xi et al., 2022), DCFSL (Zhang et al., 2022), and Gia-CFSL (Zhang et al., 2022) shown in Table 5. The highlighted results in bold included in Table 5 show that the class trees achieved the best accuracy and worst results for the class Meadows. This could be because no related word embeddings were found for the class meadows. Other classes achieved comparable classification results comparable to other models excluding the classes bitumen and shadows. This could be because of the imbalanced distribution of classes or the fluctuation caused in



(a) (b) (c)

Fig. 9 Confusion matrix for **a** Pavia U, **b** Indian Pines and **c** Kennedy Space Centre

Table 9 Comparison of OA achieved by CNN in comparison to base models

Model	PU PU		PU IP		PU KSC	
	OA	KC	OA	KC	OA	KC
Random forest	**98.34**	**0.9545**	73.9	0.75	**72.43**	**0.62**
Support vector machine	95.35	0.9052	**80.54**	**0.78**	54.16	0.5
Convolutional neural network	90.52	0.9534	75.08	0.77	67.58	0.55

the spectra of these classes due to different spatial distribution, occlusion or difference in the imaging environment. However, the highest overall classification accuracy and kappa coefficient was achieved by the proposed model.

Category 3: T_r (IP, PU) $\rightarrow$ T_e (KSC)

The third category of the experiment included testing on the KSC dataset. The last two columns in Table 6 shows the results obtained by the proposed model when trained using IP and PU. The other cross-domain models used for comparison, shown in Table 6, were tested on KSC. The results were quite interesting as the proposed model when tested using KSC achieved better results when trained using PU rather than IP. Out of 13 tested classes, KSC achieved the best results in 9 classes when trained using Pavia U. However, results were comparatively lower on the KSC dataset when trained using IP. This could be due to difference in basic attributes of two datasets despite being captured using the same sensor. This could also be because of changes in spectra due to difference in the imaging environment or inconsistent weather conditions.

Category 4: T_r (PU) $\rightarrow$ T_e (PU), T_r (IP) $\rightarrow$ T_e (IP)

The proposed model achieved excellent results when training and testing were performed on the same sets. This category was formed to test the model's performance with training and testing completing similar sets. Furthermore, the results were quite obvious. Table 7 depicts the classification accuracy for each class. PU achieved 99.49 percent, and IP achieved 93.46 overall accuracies. Figure 9 shows the confusion matrix for all three datasets. The excellent performance was achieved due to the same spectra and training–testing datasets being similar. This validates the performance of the proposed deep learning spectral-spatial architecture having great feature extraction capabilities.

Table 8 shows the parameters involved in the feature and embedding model. Figure 9 shows the confusion matrix for the three datasets when trained using Pavia University (PU), used to assess the classification accuracy of the proposed work. Table 9 shows the results obtained for T_r (PU) T_e (PU), T_r (PU) T_e (IP), and T_r (PU) T_e (KSC) for the most popular base machine learning models. This table shows the overall accuracy and kappa coefficient obtained for three respective datasets. The overall accuracy was highest when it was trained on same-domain in comparison to cross-domain datasets .

Conclusions

This study introduces a novel approach to cross-domain HSI classification by challenging the traditional paradigm of separating train and testing within the same data cube. It alleviates the problem of limited labelled samples in the HSI domain. The proposed model can be trained using images having different spectral resolutions acquired and pre-processed using different sensors. The generated word embeddings are used to associate the training and testing datasets by narrowing the gap between them. The presented feature model attempts to utilize the associations between visual elements and labelled word embeddings. Using the cosine similarity metric, the learned feature mapping is applied to testing data to determine the final label. The proposed model achieved excellent overall classification performance. However, some of the classes in testing datasets achieved low classification performance. It may be due to the similarities in the spectral distribution of some classes or to less difference in the textual description of these classes. This study proved that there is excellent scope in cross-domain classification due to the limited labelled HSI datasets available and the majority of existing models working on the same datasets. Having the strong fine spectra ability of hyperspectral images, the proposed model can be pretrained using multisource domains like military, mining, defence, or medical in the future. However, it will be interesting to see the overlap in spectra and whether it can be reduced with added word embeddings.

Funding No funding was received for conducting this study. The authors have no relevant financial or non-financial interests to disclose.

Data Availability The open data used in this research is available at https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes. The data used in this research is referenced in the source.

Declarations

Conflict of interest The authors declared that they have no conflict of interest.

Human and Animal Rights No animal was harmed during this research.

References

Akata, Z., Perronnin, F., Harchaoui, Z., & Schmid, C. (2016). Label-embedding for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(7), 1425–1438. https://doi.org/10.1109/TPAMI.2015.2487986

Alotaibi, B., & Alotaibi, M. (2020). A hybrid deep ResNet and inception model for hyperspectral image classification. *PFG—Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88*(6), 463–476. https://doi.org/10.1007/s41064-020-00124-x

Biswas, S., & Annadani, Y. (2018). Preserving semantic relations for zero-shot learning. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 7603–7612). Presented at the 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR), Salt Lake City, UT: IEEE. https://doi.org/10.1109/CVPR.2018.00793

Chi, Q., Lv, G., Zhao, G., & Dong, X. (2022). A novel knowledge distillation method for self-supervised hyperspectral image classification. *Remote Sensing, 14*(18), 4523. https://doi.org/10.3390/rs14184523

Dang, L., Weng, L., Hou, Y., Zuo, X., & Liu, Y. (2023). Double-branch feature fusion transformer for hyperspectral image classification. *Scientific Reports, 13*(1), 272. https://doi.org/10.1038/s41598-023-27472-z

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). Presented at the 2009 IEEE computer society conference on computer vision and pattern recognition workshops (CVPR Workshops), Miami, FL: IEEE. https://doi.org/10.1109/CVPR.2009.5206848

Firat, H., Asker, M. E., Bayindir, M. İ, & Hanbay, D. (2022). 3D residual spatial–spectral convolution network for hyperspectral remote sensing image classification. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-022-07933-8

Gao, K., Liu, B., Yu, X., Qin, J., Zhang, P., & Tan, X. (2020). Deep relation network for hyperspectral image few-shot classification. *Remote Sensing, 12*(6), 923. https://doi.org/10.3390/rs12060923

Graña, M., Veganzons, M., & Ayerdi, B. (2021). Computational Intelligence Group (GIC). *Hyperspectral Remote Sensing Scenes*. http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Indian_Pines. Accessed 30 Nov 2021

Guan, P., & Lam, E. Y. (2022). Cross-domain contrastive learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–13. https://doi.org/10.1109/TGRS.2022.3176637

Guo, W., Xu, G., Liu, B., & Wang, Y. (2022). Hyperspectral image classification using CNN-enhanced multi-level Haar wavelet features fusion network. *IEEE Geoscience and Remote Sensing Letters, 19*, 1–5. https://doi.org/10.1109/LGRS.2022.3167535

Lampert, C. H., Nickisch, H., & Harmeling, S. (2009). Learning to detect unseen object classes by between-class attribute transfer. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 951–958). Presented at the 2009 IEEE computer society conference on computer vision and pattern recognition workshops (CVPR Workshops), Miami, FL: IEEE. https://doi.org/10.1109/CVPR.2009.5206594

Lampert, C. H., Nickisch, H., & Harmeling, S. (2014). Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(3), 453–465. https://doi.org/10.1109/TPAMI.2013.140

Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., & Benediktsson, J. A. (2019). Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing, 57*(9), 6690–6709. https://doi.org/10.1109/TGRS.2019.2907932

Li, Z., Liu, M., Chen, Y., Xu, Y., Li, W., & Du, Q. (2022). Deep cross-domain few-shot learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1–18. https://doi.org/10.1109/TGRS.2021.3057066

Liu, B., Yu, A., Yu, X., Wang, R., Gao, K., & Guo, W. (2021a). Deep multiview learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 59*(9), 7758–7772. https://doi.org/10.1109/TGRS.2020.3034133

Liu, B., Yu, X., Yu, A., Zhang, P., Wan, G., & Wang, R. (2019). Deep few-shot learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 57*(4), 2290–2304. https://doi.org/10.1109/TGRS.2018.2872830

Liu, S., Shi, Q., & Zhang, L. (2021b). Few-shot hyperspectral image classification with unknown classes using multitask deep learning. *IEEE Transactions on Geoscience and Remote Sensing, 59*(6), 5085–5102. https://doi.org/10.1109/TGRS.2020.3018879

Lv, W., & Wang, X. (2020). Overview of hyperspectral image classification. *Journal of Sensors, 2020*, 1–13. https://doi.org/10.1155/2020/4817234

Mei, X., Pan, E., Ma, Y., Dai, X., Huang, J., Fan, F., et al. (2019). Spectral-spatial attention networks for hyperspectral image classification. *Remote Sensing, 11*(8), 963. https://doi.org/10.3390/rs11080963

Pan, E., Ma, Y., Fan, F., Mei, X., & Huang, J. (2021). Hyperspectral image classification across different datasets: A generalization to unseen categories. *Remote Sensing, 13*(9), 1672. https://doi.org/10.3390/rs13091672

Paoletti, M., Haut, J., Plaza, J., & Plaza, A. (2018). Deep&Dense convolutional neural network for hyperspectral image classification. *Remote Sensing, 10*(9), 1454. https://doi.org/10.3390/rs10091454

Radovanovic, M., Nanopoulos, A., & Ivanovic, M. (n.d.). Hubs in space: Popular nearest neighbors in high-dimensional data.

Ranjan, P., & Girdhar, A. (2022). Xcep-Dense: A novel lightweight extreme inception model for hyperspectral image classification. *International Journal of Remote Sensing, 43*(14), 5204–5230. https://doi.org/10.1080/01431161.2022.2130727

Reed, S., Akata, Z., Lee, H., & Schiele, B. (2016). Learning deep representations of fine-grained visual descriptions (pp. 49–58). Presented at the proceedings of the IEEE conference on computer vision and pattern recognition. https://openaccess.thecvf.com/content_cvpr_2016/html/Reed_Learning_Deep_Representations_CVPR_2016_paper.html. Accessed 18 Jan 2023

Shigeto, Y., Suzuki, I., Hara, K., Shimbo, M., & Matsumoto, Y. (2015). Ridge regression, hubness, and zero-shot learning. In A. Appice, P. P. Rodrigues, V. Santos Costa, C. Soares, J. Gama, & A. Jorge (Eds.), *Machine learning and knowledge discovery in databases* (Vol. 9284, pp. 135–151). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-23528-8_9

Socher, R., Ganjoo, M., Manning, C. D., & Ng, A. (2013). Zero-shot learning through cross-modal transfer. In *Advances in neural information processing systems* (Vol. 26). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2013/hash/

2d6cc4b2d139a53512fb8cbb3086ae2e-Abstract.html. Accessed 18 Jan 2023

Wang, L., Li, Y., & Lazebnik, S. (2016). Learning deep structure-preserving image-text embeddings. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 5005–5013). Presented at the 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, USA: IEEE. https://doi.org/10.1109/CVPR.2016.541

Wang, H., Wang, X., & Cheng, Y. (2023). Graph meta transfer network for heterogeneous few-shot hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing, 61*, 1–12. https://doi.org/10.1109/TGRS.2022.3233591

Wu, H., & Prasad, S. (2018). Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Transactions on Image Processing, 27*(3), 1259–1270. https://doi.org/10.1109/TIP.2017.2772836

Xi, B., Li, J., Li, Y., Song, R., Hong, D., & Chanussot, J. (2022). Few-shot learning with class-covariance metric for hyperspectral image classification. *IEEE Transactions on Image Processing, 31*, 5079–5092. https://doi.org/10.1109/TIP.2022.3192712

Xue, Z., Yu, X., Liu, B., Tan, X., & Wei, X. (2021). HResNetAM: Hierarchical residual network with attention mechanism for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 3566–3580. https://doi.org/10.1109/JSTARS.2021.3065987

Yu, C., Wang, J., Chen, Y., & Huang, M. (2019). Transfer learning with dynamic adversarial adaptation network. In *2019 IEEE international conference on data mining (ICDM)* (pp. 778–786). Presented at the 2019 IEEE international conference on data mining (ICDM), Beijing, China: IEEE. https://doi.org/10.1109/ICDM.2019.00088

Zhang, Y., Li, W., Zhang, M., Wang, S., Tao, R., & Du, Q. (2022). Graph information aggregation cross-domain few-shot learning for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2022.3185795

Zhong, Z., Li, J., Ma, L., Jiang, H., & Zhao, H. (2017). Deep residual networks for hyperspectral image classification. In *2017 IEEE international geoscience and remote sensing symposium (IGARSS)* (pp. 1824–1827). Presented at the 2017 IEEE international geoscience and remote sensing symposium (IGARSS), Fort Worth, TX: IEEE. https://doi.org/10.1109/IGARSS.2017.8127330

Zhong, Z., Li, J., Luo, Z., & Chapman, M. (2018). Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing, 56*(2), 847–858. https://doi.org/10.1109/TGRS.2017.2755542

Zhu, Y., Zhuang, F., Wang, J., Ke, G., Chen, J., Bian, J., et al. (2021). Deep subdomain adaptation network for image classification. *IEEE Transactions on Neural Networks and Learning Systems, 32*(4), 1713–1722. https://doi.org/10.1109/TNNLS.2020.2988928

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/362570474

A Fast & efficient Noise-resistant Recommender System using Semantic Ontology and User Networks.

Preprint · May 2022

DOI: 10.13140/RG.2.2.17174.22084

CITATIONS
0

READS
15

4 authors, including:

Ayush Tanwar
Delhi Technological University
5 PUBLICATIONS **3** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Folksonomy Based Fuzzy Filtering Recommender System View project

All content following this page was uploaded by Ayush Tanwar on 20 August 2022.

The user has requested enhancement of the downloaded file.

A Fast & efficient Noise-resistant Recommender System using Semantic Ontology and User Networks

Ayush Tanwar[1*], Vinod Kumar[1], Chitraksh Grover[1] and Arhan Relan[1]

[1*]Department of Computer Engineering, Delhi Technological University, Main Bawana Road, New Delhi, 110042, Delhi, India.

*Corresponding author(s). E-mail(s): ayushtw@gmail.com;
Contributing authors: vinod_k@dtu.ac.in;
chitrakshgrover@gmail.com; arhanrelan@gmail.com;

Abstract

In this era of digitalization, most domains thrive on online mediums, which provide a wide-ranging choice of resources to their users. Hence, a growing inclination towards effective recommendation systems for showing users relevant recommendations has become evident. The current shortage of semiconductor chips has led to a scarceness of computing power. The time and resources most platforms can allot to monotonous task of recommendation are limited. Moreover, each hour a considerable number of items and users are being added to platforms like Youtube, Netflix, Udemy, etc. The demand for efficient and accurate recommender systems is higher than ever. Therefore, We propose a Fast, Accurate and Efficient recommendation model: Recommender System using Collaborative Filtering, Ontology and User Networks (ReCONet), that broadens the concepts of ontology into the user domain by introducing user networks. Our approach seems uncomplicated, but extensive experimentation confirms that it performs significantly better than state-of-the-art Neural Network-based and Natural language-based approaches. Our model performed 61% faster and 18% more accurately than other state-of-the-art approaches on publicly available datasets.

Keywords: Hybrid Recommender System, Ontology, User Networks, Content based system, Collaborative Filtering

1 Introduction

The advent of the internet has led to an exponential increase in the amount of information available online. This overfeeding of information to the users has been countered with various information retrieval and informational Filtering systems. Recommender Systems are a popular example of such systems. They are used to predict a set of items for a user based on her/his preference. These have become very valuable for online businesses as they help customers in finding the desired product. According to Gomez-Uribe and Hunt [1] users on Netflix are not able to articulate their viewing preferences and lose interest after 60 to 90 seconds of choosing and reviewing 10-20 titles. Thus recommendation poses a great business value since 80 % of the hours streamed on Netflix are from the recommended items. Recommender Systems[2] have been extensively used in movies[1][3], music [4], Social Networks[5] etc.

There are two broad categories of Recommender systems: Content-based Recommender systems and Collaborative Filtering Recommender systems [9] [10] [5] [11]. Content-based Recommender systems[12] uses item profiles to recommend similar items. However it can lead to only a few items being recommended. Collaborative Filtering being the more popular approach uses the user item rating matrix as user's feedback for previously selected items. This matrix is used to find similar users and items and predict a user item rating.

Collaborative Filtering systems are further divided into memory based algorithms and model based algorithms. Memory based algorithms utilises the complete user item rating matrix to find the neighbourhood of a user(user based) or similar items(item based) to predict the ratings. Since they requires the complete rating matrix for prediction , they are computationally expensive and have scalability issues. In contrast, Model based algorithms use clustering algorithms to determine the neighbourhood of a particular user and provide recommendation based on that cluster. These algorithms do not have large memory requirements and scale well for large databases. However the quality of the prediction is low as compared to their counterparts.

Other drawbacks of Collaborative Filtering recommender systems include sparsity in the rating matrix and cold start problem. Various dimensionality reduction techniques[13][11] have been used to deal with sparsity. For the cold start problem[14], recommender systems use various knowledge based techniques which aid in the recommendation process. Ontology[15][13] is a knowledge based approach widely used in the recommender systems. Ontology is defined as conceptualization of a domain. In Ontology, items are linked to other items in the domain based on their properties, which is then used to find items similar to the item under consideration for recommendation. Another widely used knowledge-based approach uses social networks and trust relationships[16][17][18] of a user for finding its neighbourhood and then recommending items. These approaches have their own shortcomings. Creating ontology and using these relations to find similar items is not a trivial task.

Moreover, using social relations to find the neighborhood of a user in a recommendation perspective may not be effective since a user may have a different taste and needs as compared to his/her friends/acquaintances.

We propose a hybrid Collaborative Filtering recommender system, Re-CONet, to provide personalised recommendations to a user. The hybrid approach utilises memory-based and model-based algorithms along with user and item based approaches to counter the drawbacks of each approach. We further use item ontology and introduce user ontology as user networks to enhance user-item profiling, assisting the recommendation process. The major contributions of our study are as follows:

- **Introducing User Ontology as User Networks** - User networks are used to find similarity between the users based on their social relations. Contradictory to previous approaches, user networks are used on top of the Collaborative Filtering framework for rating prediction.
- **Detailed Item Ontology** - We perform an in-depth analysis of our item domain to create a vivid ontology that utilises the information content given by different item features taking in consideration their conceptual dissimilarity.
- **Density based clustering approach** - We use DBSCAN algorithm for user and item clustering based on their rating patterns. Since it is an adaptive algorithm, it forms organic clusters to be used for Collaborative Filtering and the outliers are separated as noise.
- **Simple and intuitive approach** - The primary aim is to provide personalised recommendations to users even for the outliers. The hybrid approach compounds the benefits of each of its constituents to identify and accommodate the specific needs of each user.

The remainder of this paper is organized as follows: In Section 2 we present the previous approaches used in the field of recommender systems. In Section 3 we present the Re-CONet model. Section 4 presents experimentation and results and Section 5 concludes our study.

2 Related Work

Before presenting our hybrid approach and experimental results , we give a brief overview of previously used approaches, methods and algorithms. Recommendation of relevant items can be given on the basis of similarity between items or users (Content-based), similar rating patterns of a group of users (Collaborative Filtering) or a hybrid approach. The type of approach usually depends on the information, metadata and feedback available. Content-based approach is used when sufficient information and semantic relations are available beforehand whereas Collaborative Filtering is useful for a large user base. Generally the approach is formed according to the experience, expertise and intuitively for a particular domain. Our approach is a hybrid approach , which uses a combination of various recommendation techniques to deliver the best recommendation.

2.1 Collaborative Filtering

Collaborative Filtering techniques [5] are used for providing customized recommendations by calculating the similarity between different users with similar choices. Collaborative Filtering can be classified majorly into two categories: Memory Based Approaches and Model Based Approaches.

Memory Based Collaborative Filtering uses all the user item data available . They base their recommendation on the fact that similar users are more likely to have similar tastes, so their recommendations should also be similar. It is a community goaded technique. Memory Based Recommendation[19] can be further divided into Item based technique and User based technique. In User Based technique, users who rated the items similarly may have same taste. Based on this , it recommends new items which are rated highly by similar users. Item Based technique focuses on items , it assumes that items which are rated similarly are presumably similar and therefore it recommends items with highest similarity.

Model Based Collaborative Filtering (Tommaso Di Noia et al[20]) (Hossein Javedani Sadaei et al[21]) gives item recommendation using a pre-constructed model. The model is built by various ML techniques such as Bayesian networks, clustering, etc. The major drawback of Collaborative Filtering is cold start problem, This problem occurs due to the insufficiency of data for some users or items[14].

2.2 Ontology

Ontology is originally defined by Thomas R. Gruber[22] as "formal specification of a shared conceptualization". Ontology is a formulation of a particular field like movies, courses, construction materials, etc into a human-apprehensible, but computer understandable form consisting of entities and attributes along with their various relationships and axioms.

Over the years many recommendation approaches involved combining Collaborative Filtering techniques with ontology to improve the prediction accuracy. Daramola et al[23] implemented an online tourism recommender system for tourism related services. Ontology enables and shares the structure of information among people and software agents (Ahmad Hawalah et al)[24].

Ontology in recommender systems virtually remedies the cold-start problem. Stuart E. Middleton et al[15] experimented with ontology based Recommender Systems to observe the effects of ontology on cold start problem. They tested their Quickstep recommender system for 7 weeks , they started with an empty log to simulate the conditions of cold-start problem and applied bootstrapping algorithms . After week 7 they examined and concluded that the cold start problem was completely eliminated.

2.3 Trust

Trust network between different users connects them to enhance the prediction accuracy and in turn enhance the final recommendation. Trust Networks are

built on the basis that a user 'a' who trusts other user 'b' is most likely interested in items liked or bought by 'b'. Trust can be implemented in two ways – probabilistic and gradual.

Probabilistic implementation (Y. Wang [17],Y. Kim [18] and A. Abdul Rahman[25]) is engaged with a trust factor which is measured by a '0' or a '1' which means a user 'a' either trusts a user 'b' or user 'a' does not trust user 'b'.

Gradual implementation (W. Li[26] ,J. Wu[27], C. Chin Chen[28]) on the other hand does not deal with absolute '0' or '1' , it deals with the extent to which a value is zero or the extent to which a value is 1. So this means in gradual implementation we find the extent to which user 'a' trusts user 'b'. This approach is more fitting in real life scenarios. It fits Fuzzy logic.

Few properties of trust in previous approaches (Golbeck 2017[29]) are –

- asymmetric: X trusts Z does not imply that Z trusts X.
- distributive: X trusts Y and Z, does not imply that Y trusts X and Y trusts Z.
- transitivity: While some of the approaches consider trust as non-transitive such as Abdul Rahman et al. (2000)[25], most of the works consider transitivity as an important property of Recommender system like Weijiang Li (2017)[26], Gohari (2017)[30], Josang and Pope (2005)[31] .

2.4 Hybrid Recommender Systems

Hybrid approaches combines two or more approaches to improve final recommendation. Balabanovic and Shoham [32] were one of the first to implement a hybrid approach for recommender systems. Their Fab system recommended web pages by combining implicit and explicit assistance.

Barragáns-Martínez and Costa-Montenegro [33] proposed a hybrid TV program recommender system. It used a Content-based and an item based Collaborative Filtering approach. It automatically maps user's tastes with a television program.

Neural Network based approaches[6][8] are common these days. Some of the neural network based approaches are described in brief in this section. We have also experimented and analysed the prediction accuracy and computation times of these approaches against our own approach in Experimentations section.

The DNNRec[6] utilizes embeddings for describing items and users to learn non-linear latent factors. Their approach ameliorates the cold start problem by incorporating side information regarding users and items into a deep-learning neural network. Their approach uses the embeddings to create Neural network to improve accuracy further.

Paradarami et al. [8] used a set of collaborative and content based features for a hybrid neural network based approach to counter standalone memory approach. Walek et al. [3] introduced a hybrid monolithic RS named Predictory. Predictory combines collaborative filtering techniques, content-based

techniques, and the fuzzy expert technique. Predictory's final n recommendations were computed using a fuzzy expert system, which evaluates the significance of the movies. The expert system worked with numerous parameters – the number of ratings, the similarity measure between already rated movies, average movie rating, etc.

Kamath et al.[7] developed a smart news recommendation search engine that recognizes similar news articles available from numerous sources on a particular day, then applies a semantics-based enhanced bag-of-words technique to leverage the efficiency of search recommendation. Natural-language-processing methods were used to identify articles dealing with similar news and serve on a user query with summarised news from various sources.

Hybrid Recommender Systems showed improved results in comparison to conventional Collaborative Filtering based recommendation systems.

3 Re-CONet Model

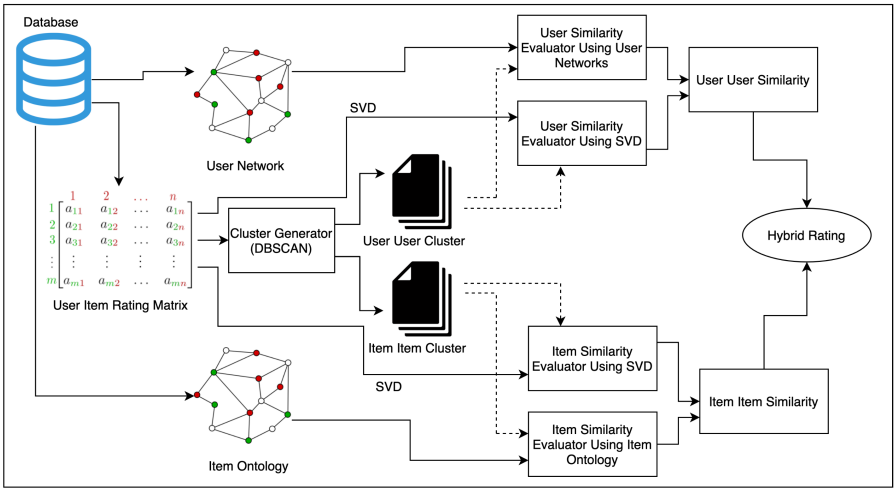


Fig. 1 Graphical representation of the proposed Re-CONet architecture

In this section we present our hybrid approach for the recommender system that uses ontology along with Collaborative Filtering for predicting ratings and providing recommendations.

In a traditional Collaborative Filtering system , users' preference regarding the items are taken into consideration for giving recommendations to the user. These preferences are generally stored in the form of a user-item rating rating matrix. Based on these ratings, users having 'similar liking' are identified and items are recommended.

In order to find similarity between the users or items and make predictions, it is required that the complete rating matrix be present in the memory. Such

Table 1 Rating Matrix

User/Item	$\mathbf{I_1}$	$\mathbf{I_2}$	...	$\mathbf{I_j}$	...	$\mathbf{I_n}$
$\mathbf{U_1}$	R_{11}	R_{12}	...	R_{1j}	...	R_{1n}
$\mathbf{U_2}$	R_{21}	R_{22}	...	R_{2j}	...	R_{2n}
...	...	...	...	...	...	...
$\mathbf{U_i}$	R_{i1}	R_{i2}	...	R_{ij}	...	R_{in}
...	...	...	...	...	...	...
$\mathbf{U_m}$	R_{m1}	R_{m2}	...	R_{mj}	...	R_{mn}

type of systems called memory-based systems produce highly effective ratings, although their memory requirements are very high. Another type of Collaborative Filtering systems are model-based , where a model is trained offline and the predictions made online using the same. This makes the process of recommendation considerably fast, however the predictions are not as accurate as given by their counterparts. Our proposed model implements the best of both approaches in a way that makes the recommendation process efficient and accurate.

Figure 1 shows the architecture of our model. The model considers two types of databases. One consist of the user-item rating matrix and the other contains the ontologies. The rating matrix is defined as follows.

Rating Matrix: It is the matrix containing ratings of N items by M users where row headers represent the users and the column headers represent the items as shown in *Table 1*.

Let U_i be the i^{th} user and I_j be the j^{th} item , then R_{ij} is the rating given to I_j by U_i. $R_{ij} \in [0,5]$. An item not rated by a user is represented by a default value 0.

Using these ratings, users and items are clustered into different clusters. This is done in offline phase. The clusters learned establishes the neighbourhood of a user/item. The neighbourhood is then used while computing the similarity measures and providing recommendations online. Since the neighbourhood restricts the number of users/items to be considered , it makes the process considerably efficient.

Since a user rates only a few items , this makes the rating matrix highly sparse. The similarity between users/items becomes skewed since all the unrated items are given default rating 0. Thus, before calculating similarity for different users in a cluster we use dimensionality reduction techniques to reduce the latent factors , which decreases the sparsity in the matrix. Based on these similarities , a rating R_{ij} is predicted by aggregating the ratings provided to item I_j by the users in the U_i's cluster or vice-versa in case of item-item clustering.

Method described above uses rating patterns to find similarity between users or items and give recommendations. A rating is a user's view on an item and thus using rating patterns for recommending an item to a user we exploit the fact that users having similar rating patterns have similar kind of interests and therefore an item liked by a user might also be liked by the other user. Similarly similarity between items is also calculated based on how they

are rated. Apart from the user's outlook on an item, similarity between items and users can also be calculated on the basis of their characteristics. This kind of approach uses the fact that items having common features are similar. Considering movies as items and genre as one of its characteristics , a user interested in a movie belonging to a particular genre may also be interested in other movies of the same genre. Ontology is a collection of concepts and categories in a subject area or a particular field (Eg.- Movies) which are used to represent their properties and the relationships between them. Hence all the instances of that particular domain can be mapped into corresponding concepts and categories , thus relating the instances. The similarity between instances using such kind of model is called Ontological Similarity.

In our model considering two domains - users and items , the ontology can be classified into:

- Item Ontology
- User Ontology

3.1 Cluster generator

It is a learning step and is done offline. Here the rating matrix is taken and clusters are formed that consist of neighborhood of users that resembles in terms of their likes-dislikes and that of items , in terms of their likeliness. Two kinds of clusters are formed.

User-user clusters : users are clustered together into different clusters based on how they rate different items. Each row in the matrix is considered as a user vector of that particular user. It is defined as follows Let U_k be a user vector, Then :

$$U_k = r_{kj}$$

$$r_{kj} \textit{ is the rating of } j_{th} \textit{ item by } k_{th} \textit{ user } \forall\, j \in (1, \mathsf{Items}) \,, k \,\epsilon\, (1,\ \mathsf{Users})$$

Therefore, a user vector consists of the ratings provided by the user to all the items. Hence it is a data point in N-dimensional data space, where N is the total number of items . These data points are then clustered in order to form user-user clusters.

Item-Item clusters: Items are clustered together into different clusters based on how they are rated by different users. Each column in the matrix is considered as an item vector of that particular item. It is defined as follows

$$I_k = r_{jk}$$

$$r_{jk} \textit{ is the rating of } k_{th} \textit{ item by } j_{th} \textit{ user } \forall\, j \in (1, \mathsf{Users}) \,, k \,\epsilon\, (1,\ \mathsf{Items})$$

Therefore, n-item vector consists of the ratings provided to the item by all the users. Hence it is a data point in M-dimensional data space, where M is the total number of users . These data points are then clustered in order to form Item-Item clusters.

Clustering Algorithm: In our approach, we use DBSCAN [37] algorithm for clustering. It is a density based clustering algorithm , that creates clusters of high density separated by clusters of low density. It is a non-parametric algorithm in the sense that it does not pre-decide the number of clusters.

DBSCAN algorithm uses an iterative process to cluster the data and is able to handle large datasets. The advantage of using DBSCAN is that it effectively handles outliers lying in the low density regions. These outliers could be movies or users that have very few ratings and are not considered as a part of any cluster. This gives us the advantage that , instead of finding a similar pattern of rating which does not exist , we can given recommendations based only on the ontology based approach.

3.2 Cluster Similarity

One of the major drawbacks of using the conventional Collaborative Filtering is the sparsity of the rating matrix. One of the counter measures to solve the problem is to use the mean value given by the other users which provides a pseudo-negative or pseudo-positive outlook of the movie. Another way is to use dimensionality reduction technique to remove the redundancy of noisy elements. Since the aim is to have users having a similar rating pattern despite having only a few ratings to get an improved similarity, our approach combines both the methods to give better similarity results. We use SVD[40] as the dimensionality reduction technique.
SVD is based on ternary factorization of a real-valued matrix as shown

$$R_{\mathrm{mxn}} = M_{\mathrm{mxm}} \times K_{\mathrm{mxn}} \times S_{\mathrm{nxn}}, \tag{1}$$

Where M and S are orthogonal matrices [$\mathrm{MM_T} = \mathrm{I}$] and K is a diagonal matrix with elements $\mathrm{a_0} > \mathrm{a_1} > \mathrm{a_2} \ldots \mathrm{a_n}$.

The matrix K comprises of latent factors of the items which represent their various characteristics such as genre, date of release in case of movie recommendation; weather, geography in case of tourism recommendation. Thus, the matrix M represents the latent factors implicitly rated by each user. Similarly, the matrix S represents the latent factors implicitly rated for each movie.

All the latent factors exceeding a threshold are considered as necessary and those less than the threshold are seen as noise, hence they are removed.In our approach we take the threshold as 'l' .We retain only the top l x l sub-matrix to form the new diagonal matrix $\mathrm{K_l}$. According to the diagonal matrix $\mathrm{K_l}$, we simplify matrix M and S into matrix $\mathrm{M_l}$ and $\mathrm{S_l}$ and obtain the new matrix $\mathrm{R_l}$.

$$R_{\mathrm{l}} = M_{\mathrm{l}} \times K_{\mathrm{l}} \times S_{\mathrm{l}} \tag{2}$$

This is the main application of SVD for the linear approximation of the original matrix R. Research shows that removing the small singular values "noise" indeed enhances the consumer-item relationship.

Table 2 User-User Similarity

	User U_1	**User U_2**	**User U_3**	**User U_4**
User U_1	S_{11}	S_{12}	S_{13}	S_{14}
User U_2	S_{21}	S_{22}	S_{23}	S_{24}
User U_3	S_{31}	S_{32}	S_{33}	S_{34}
User U_4	S_{41}	S_{42}	S_{43}	S_{44}

In our approach, we take the average rating for every user i as r_u and replace every non-zero rating r_{ui} with r_{ui}-r_u. Then we perform the SVD on the rating matrix R to get the decomposed matrix R_l The predicted rating matrix PR is obtained using the formula:

$$PR_{ui} = r_u + R_{lui} \tag{3}$$

After generating the predicted rating matrix , similarity for each user/item in a cluster is calculated as follows. Let U be a user cluster as shown in the Table 2 , where User U_i is the i^{th} user of nth cluster and S_{ij} = Similarity between i^{th} and j^{th} user. The same procedure is carried out for item-item similarity. The similarity is calculated using Tanimoto Coefficient[38].

The Jaccard Coefficient and cosine similarity are generalised to obtain the Tanimoto coefficient where a real-valued vector of attributes represents each item/user. The Tanimoto Coefficient is computed as:

$$T(U_i, U_k) = \frac{U_i.U_k}{\|U_i^2\| + \|U_k^2\| - U_i.U_k} \tag{4}$$

where U_i and U_k are real-valued vectors representing item/user.

The final SVD Similarity, Θ_{SVD}, calculated by the Tanimoto coefficient is used for hybrid rating prediction in section 3.5 .

3.3 Item Ontology

Since constructing an ontology is domain-specific, in this section we define our Item ontology considering the movie domain. The Figure 2 represents our movie ontology graph.

The ontology graph can be broadly divided into 2 sections:

- Concept section
- Instance section

The concept section consists of movie attributes like year of release, language, genre etc arranged in a logically ordered tree with movie as the root node and genre and attributes as its child nodes. Attributes are further classified into "people involved" (eg.director, music composer etc.) and "movie properties" (eg.language, year of release etc.). Movie genres are interrelated and thus, can not be placed equidistant in the genre sub-tree. For instance, movie X having

genre Horror is more closely related to movie Y having genre Thriller than to a movie Z of genre Action. Since, genres are intuitively related, research done by cinematic professionals can be summarized into a genre sub-tree given in Figure 3.

The instance section consists of movie linked with corresponding attribute instances. For example, the movie "3 Idiots" can be linked with language "Hindi", year of release "2006". The concept and instance section are connected by mapping the corresponding attribute instance in the instance section and the attribute in the concept section.

As shown in the graph,the attributes are mapped with the corresponding attribute instances whereas the genres are mapped directly with the movie. Moreover, genres are interrelated but other movie attributes are not. Thus, genre is considered as a separate movie attribute in our movie ontology.

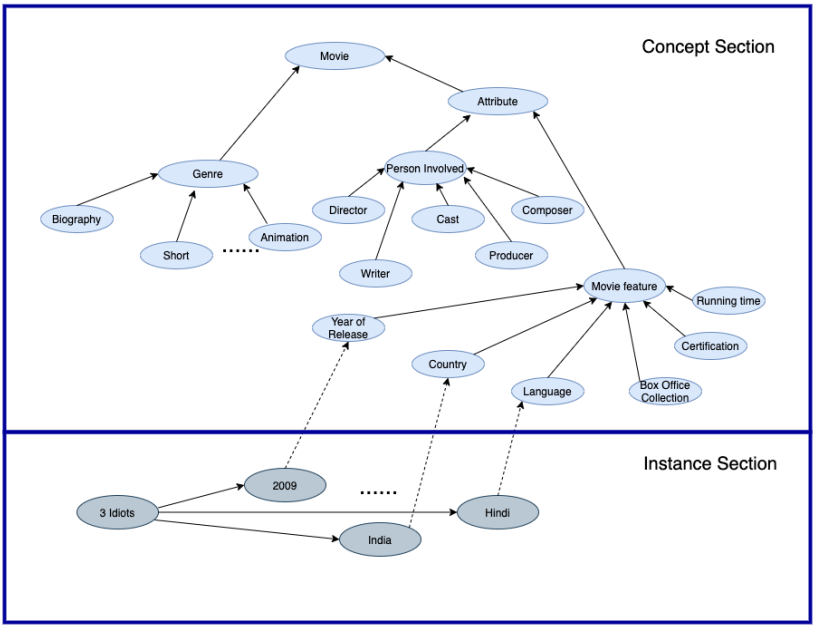


Fig. 2 Movie Ontology defined for the Re-CONet model

In our model, considering Genre and Attribute subtrees, two types of similarity measures are used in order to calculate similarities between movies:

- Genre Similarity
- Attribute Similarity

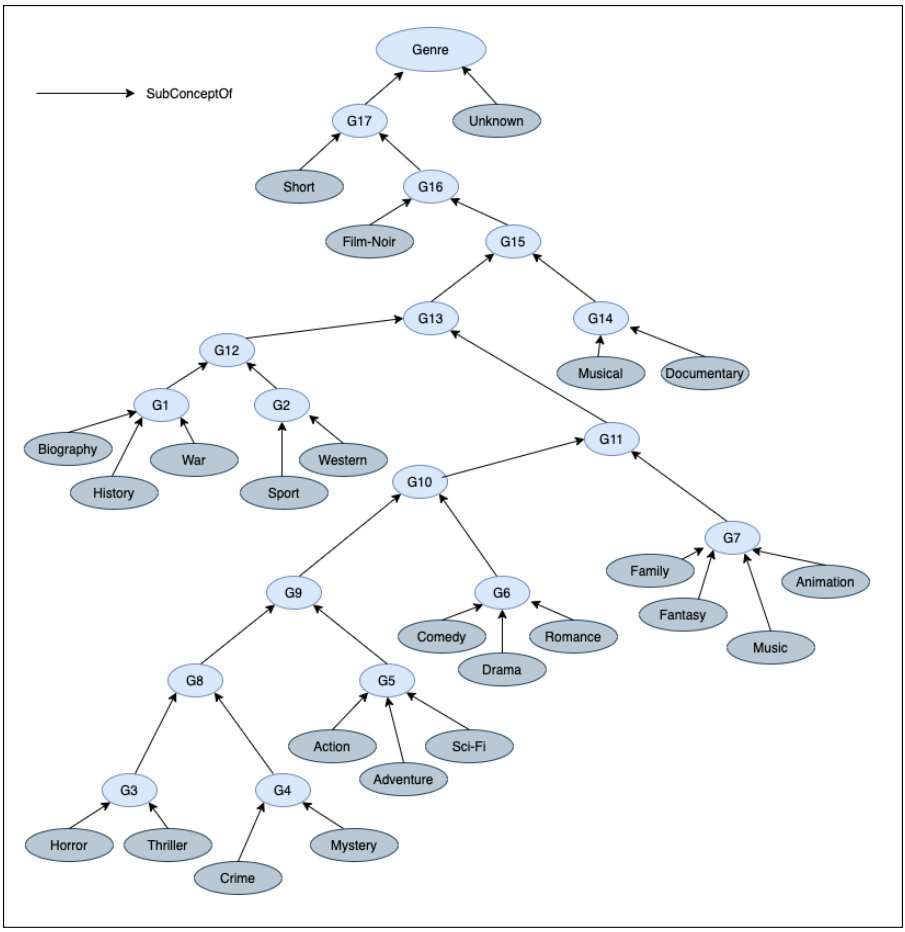


Fig. 3 Genre Ontology

For each pair of items, our model computes an ontological similarity using the weighted mean of these similarity measures.Let $\Theta_{Onto}(I_{\mathrm{i}}, I_{\mathrm{j}})$, ontological similarity between instance $\mathrm{I_i}$ and instance $\mathrm{I_j}$, be computed as :

$$\Theta_{Onto}(I_1, I_2) = \frac{x \times GS(I_1, I_2) + y \times AS(I_1, I_2)}{x + y} \tag{5}$$

Genre similarity between two movie instances (GS) is based on the shortest distance between the corresponding genres in the taxonomy. For instance, let Movie A be a comedy, Movie B a drama and Movie C an adventure movie. In our Genre Ontology, we expect that genre similarity between Movie A and Movie B is greater than the genre similarity between Movie A and Movie C.

Therefore, at the beginning, conceptual similarity between two genres (CSG) is specified. Similarity between two movies is computed using the CSG between respective genres of these movies.

Conceptual similarity between genres is calculated by an information theoretic approach [39] based on probabilistic model and is given by :

$$CSG(G_i, G_j) = \frac{log(Pr(G_k))^2}{log(Pr(G_i).Pr(G_j))} \tag{6}$$

$Pr(G_n)$ is the probability that a randomly selected movie has genre G_n, and G_k is the nearest genre that subsumes both G_i and G_j. $Pr(G_n)$ is calculated as -

$$Pr(G_n) = \frac{n(G_n)}{N} \tag{7}$$

Where N represents the total number of movie instances and $n(G_n)$ represents the number of movie instances connected to each genre subsumed by genre G_n

For instance, to calculate the similarity between Biography and War genres, let the total number of movie instances, N, be 1000 in the system. n(Biography) be 98 and n(War) is 102. The nearest genre that subsumes both Biography and War is G_1. Additionally, $n(G_1)$ be 250. Using these values CSG(Biography,War) can be computed as.

$$\begin{aligned} CSG(Biography, War) &= log(Pr(G_k))^2/(log(Pr(G_i).Pr(G_j))) \\ &= log(250/1000)^2/(log((98/1000).(102/1000))) \\ &= 1.204/2 \\ &= 0.602 \end{aligned} \tag{8}$$

GS between movie M_i and M_j is then defined as follows.

$$GS(M_i, M_j) = OSS(GSET(M_i), GSET(M_j)) \tag{9}$$

Where $OSS(S_1, S_2)$ is the similarity between sets S_1 and S_2 and $GSET(M_i)$ is the set of genres linked with the movie M_i,defined in the following sections.

Attribute Similarity

Attribute similarity between two instances (AS) is calculated using the common valued features like director, year of release, language etc.. For instance,let the Director of Movie X and Movie Y be A and director of Movie Z be B. Thus, Movie X and Y are more similar than Movie X and Z(where X and Z has no similar attributes) because of the common attribute Director.

AS between instance I_i and instance I_j is computed as:

$$AS(I_i, I_j) = 1 \;\; if \; I_i = I_j \tag{10}$$

$$AS(I_i, I_j) = \frac{\sum_k asi(I_i, I_j, k)}{N} \;\; if \; I_i \neq I_j \tag{11}$$

N is the number of links terminating at the concepts subsumed by instance I_i and I_j respectively. $ASI(I_i, I_j, k)$ gives the similarity for link k between

instances I_i and I_j. Thus, the mean of the similarities for each link of these instances gives their attribute similarity.

ASI(I_i, I_j, k) can be computed using the associated instances of I_i and I_j corresponding to the link k. For example, the instances associated for 'Has Director' link between two movie instances in our movie ontology are given by all the movies directed by these directors.

Thus, ASI(I_i, I_j, k) is calculated using the similarity between the sets of their associated instances.

ASI(I_i, I_j, k) is computed as.

$$ASI(I_i, I_j, k) = 0 \;\; if \; As(k, I_i) = \phi \; \forall \; As(k, I_j) = \phi \tag{12}$$

$$ASI(I_i, I_j, k) = CSG(As(k, I_i), As(k, I_j)) \;\; otherwise \tag{13}$$

Where As is the set of instances associated with the instance with respect to link k

Ontological Similarity Between Sets (OSS)

Similarity between two sets is required in order to calculate genre similarity and attribute similarity. It relies on the similarities between their elements. Genres for genre similarity and instances for attribute similarity calculation are the elements of the sets. Similarity between sets S_1 and S_2 (OSS(S_1, S_2)) is the following.

$$OSS(S1, S2) \; = \; \frac{(\mid S1 \wedge S2 \mid + \sum_{(I,J)} CSG(I, J))}{(\mid S1 \wedge S2 \mid + \mid S1 - (S1 \wedge S2) \mid . \mid S2 - (S1 \wedge S2) \mid)} \tag{14}$$

$$where \; I \in S1, J \in S2, (I, J) \in (S1 \vee S2) - (S1 \wedge S2)$$

OSS is used to calculate the ontological similarity, Θ_{Onto}, which is further combined with item-based SVD Similarity, Θ_{SVD}, as shown in section 3.5 .

3.4 User Networks

In Item Ontology, we considered the links between items and their features and calculated the similarity between items. This method is not feasible for users as they have innumerable features like age, gender, religion, ethnicity, etc. Instead of considering each feature explicitly, we exploit the social links between the users. For instance, friends on a social networking site share various common interests like the institution they work in or the area they live in. So rather than using features like institution or area to calculate similarity, we directly exploit their social links. Trust networks, which have been widely used in Recommender systems, is also an example of User Ontology.

The previously used approaches tend to recommend similar items among users who are connected, which is not always beneficial as the preferences between two connected users may vary in some cases. Since users belonging to

the same cluster have the same preferences, our approach uses user network to find similarity between users based on their cluster.

A graph between these users is created that represents our user network. Figure 4 is a part of user network graph.

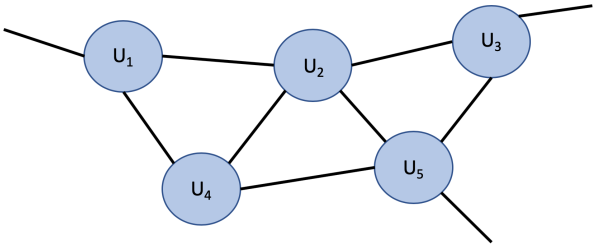


Fig. 4 User Network graph illustrating a user ontology. The nodes represent the users with connecting edges as the social connections between them.

To calculate the similarity between any two users in the system using the User network we use the concept of friend of a friend. Assuming that user U_i is connected to user U_j and further user U_j is connected to user U_k, then :

$$\Theta_{Net}(U_i, U_j) = 1 \; and \; \Theta_{Net}(U_j, U_k) = 1$$

$$0 < \Theta_{Net}(U_i, U_k) < 1$$

Where 'Θ_{Net}' is the similarity between two users. Using this concept we calculate the similarity between each user in a cluster as:

$$\Theta_{Net}(U_i, U_j) = \frac{1}{d(U_i, U_j)^p} \tag{15}$$

Where $d(U_i,U_j)$ is the distance between users U_i and U_j in the user network graph. p is a parameter that determines the gradual decrease in the similarity value . Generally $0 < p < 1$. For $p = 1$ similarity is inversely proportional to the distance between users.

3.5 Rating Prediction

Similarities computed in the Similarity calculation layer are then used to give a final Hybrid Recommendation.

Firstly, Similarity from Item-based CF with SVD and Item-Ontology are linearly combined to give item rating as shown below

$$Item\ Rating = b_{ij} + \frac{\sum_k (\alpha_{item} \times \Theta_{SVD} \times r_{kj} + (1 - \alpha_{item}) \times \Theta_{Onto} \times r_{kj})}{\sum_k (\alpha_{item} \times \Theta_{SVD} + (1 - \alpha_{item}) \times \Theta_{Onto})} \tag{16}$$

α_{item} is Hybrid Linear Combination Factor for combining similarity from Ontology and Collaborative Filtering r_{kj} is rating of j^{th} item by k^{th} user b_{ij} is global standard rating of j^{th} item by i^{th} user Similarly the User-based CF with SVD and User-Network similarity are combined to give user rating as shown below

$$User\ Rating = b_{ij} + \frac{\sum_k(\alpha_{user} \times \Theta_{SVD} \times r_{ik} + (1 - \alpha_{user}) \times \Theta_{Net} \times r_{ik})}{\sum_k(\alpha_{user} \times \Theta_{SVD} + (1 - \alpha_{user}) \times \Theta_{Net})} \tag{17}$$

α_{user} is Hybrid Linear Combination Factor for combining similarity from Ontology and Collaborative Filtering r_{ik} is rating of i^{th} user and k^{th} item b_{ij} is global standard rating of j^{th} item by i^{th} user

Finally, we calculate the rating prediction by combining the User rating prediction with Item rating prediction by using the formula

$$Hybrid\ Rating = \beta \times (Item\ Rating) + (1 - \beta) \times (User\ Rating) \tag{18}$$

β is Weighting factor for combining ratings according to their respective weights.

4 Experimentation and Analysis

4.1 Data Exploration

For the evaluation of our model, we have considered the Flixster dataset [34] & MovieLens 20M dataset [35]. We also extracted attributes and reviews for movies using IMDBPy[36] web crawler. Flixster is a social movie site allowing users to share movie ratings, discover new movies and meet others with similar movie preferences. The dataset was cleaned prior to using according to our requirements.The Flixster dataset is a popularly used dataset in the recommendation domain consisting of 2,523,386 users and 9,197,338 links among them according to their shared preferences. The movie ratings given by these users are on a scale of 0-5 and stored in a user item rating matrix. MovieLens 20M dataset has 20 million movie ratings along with 465,000 tags applied to 27K movies by about 138K users. We have combined the Data from Flixster and MovieLens 20M DataSet and for constructing the movie ontology. We randomly split the data into 80% for training and 20% for testing purposes. We have performed all the experimentation on a system with 16 GB of ram, 3.2 Ghz i5 Quad core processor and GPU GTX-1060 with 6GB dedicated video memory.

The Figure 5 and 6 depict the clustering on the basis of rating patterns of different users and items. Since the rating patterns vary on the basis of multiple factors, they are sampled along 10 major principal components. We obtained around 50 clusters each for users and items and have shown 6 clusters to show the similarity in rating pattern. As it can be seen that items and users in the

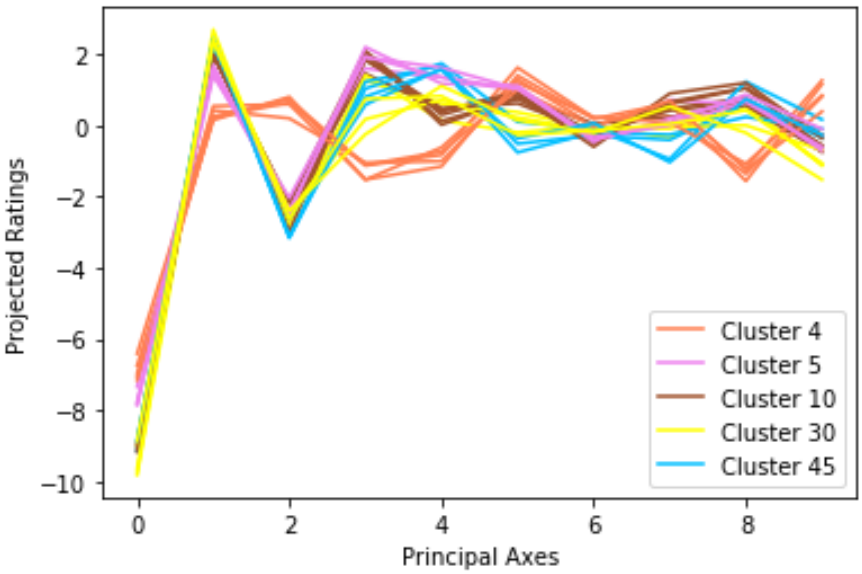


Fig. 5 Rating Patterns of Different User Cluster

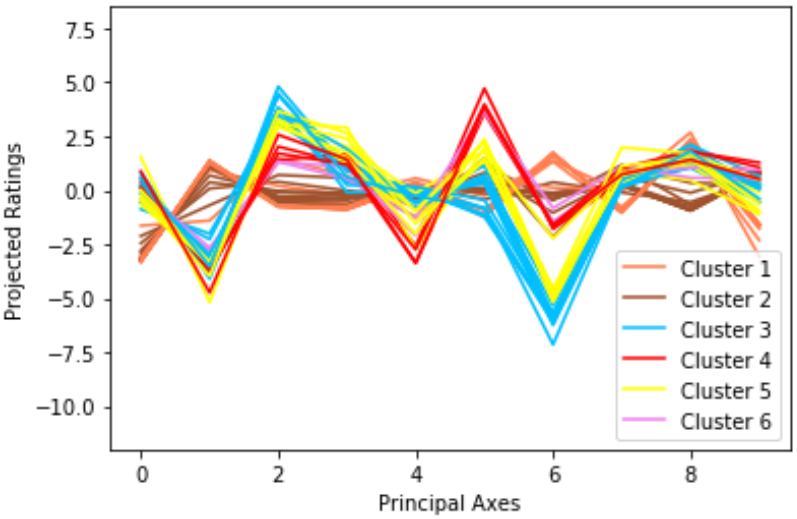


Fig. 6 Rating Patterns of Different Item Cluster

same cluster deviate only at a few points in the rating trajectory, they can be said to have similar yet not same rating patterns. Thus, the required rating is predicted by considering the corresponding average rating given by the users or for the items in the same cluster.

Table 3 shows the variation of sparsity in rating matrix with the number of chosen latent factors. As we can see, the sparsity shows a gradual decrease

Table 3 Sparsity vs latent factors

Number of latent factors (K)	Percentage of sparsity
Without SVD	99.33
10	0.360
50	0.330
75	0.330
100	0.330
500	0.310
1000	0.300
2000	0.230
3000	0.210
4000	0.210
5000	0.018

with increase in the number of chosen latent factors and drastically drops as the number of latent factors increases to 5000 which is chosen as the required value. Thus, SVD aids in reducing the sparsity of the rating matrix.

One of the fundamental criteria for a good recommender system is the personalisation of recommendations provided to each user. In this respect, to evaluate our model we calculate the similarity between the recommendations provided to users belonging to different clusters. The mean inter-cluster similarity is calculated as :

$$\overline{ICS} = \frac{\sum_m \sum_n (R_m \wedge R_n)/(R_m \vee R_n)}{\mid C_i \times C_j \mid} \tag{19}$$

where C_i and C_j are the sets of items in cluster i and j respectively, R_m and R_n are the sets of items recommended to user m and n belonging to clusters C_i and C_j respectively.

The 2-D plot illustrated by Figure 7 shows the $\overline{ICS}$ between 10 random clusters with low similarity values represented by light areas and dark areas corresponding to high similarity values. As we can see, distinct clusters show considerably less similarity as compared to their intra-cluster similarity. Thus, the primary aim to provide personalised recommendations to each user is fulfilled by our model.

4.2 Evaluating Re-CONet Model

The proposed hybrid Re-CONet model is evaluated and compared with other state-of-the-art recommender system models.

We compared the results of the Re-CONet model along with 4 other models for different values of noise percentage. Noise refers to items having irregular rating patterns and thus, do not belong to any cluster. Given the size of the sample dataset used for evaluations, we randomly chose the noisy items and clustered items according to the percentage of noise required.Since the dataset in consideration is quite large, the number of relevant items may seem negligible in some cases and thus MAE and RMSE may mislead sometimes. In order

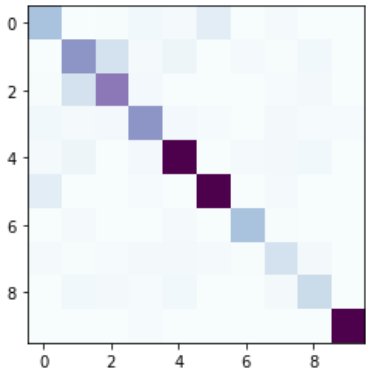


Fig. 7 Similarity between recommendations provided to users belonging to different clusters.

Table 4 Evaluation and comparison of Re-CONet Model with various approaches at different noise levels.

Approach	F-Score (20% Noise)	F-Score (30% Noise)	F-Score (40% Noise)
Paradarami [8]	0.621	0.572	0.547
Predictory [3]	0.689	0.673	0.591
Kamath [7]	0.694	0.658	0.612
DNNRec [6]	0.721	0.712	0.627
Our Approach	**0.756**	**0.743**	**0.741**

to solve such imbalanced classification problems, decision support metrics such as Precision and Recall are used. The precision evaluates the relevance of the recommended items. On the other side, the recall evaluates the tendency to recommend all the relevant items.

A balance between recall and precision is obtained by computing the F score as the harmonic mean of precision and recall. In our approach we have used F score to compare various models. It is given by the formula:

$$F = \frac{2.Recall.Precision}{Precision + Recall} \tag{20}$$

We have compared our model with 2 exceptional neural network based approaches Paradarami RS [8] and DNNRec [6], a natural language based approach Kamath RS [7] modified for recommending movies and a fuzzy expert system based approach Predictory[3]. In order to evaluate the Re-CONet model using decision-support metrics, we considered different percentages of noise (in increments of 10%). Table 4 shows the F-Score values for various models. From the table we can see that the F-Scores obtained by the Re-CONet model is relatively high in comparison to other models. The table also shows the change in F-Score with change in noise.

Paradarami [8] used a group of collaborative and content features along with reviews metadata to develop a neural network model. The approach was

created to reduce rating misclassification error, and it used the stochastic gradient descent optimization algorithm to do the same. The approach seemed sound in theory, but in practice, due to the black-box nature of neural networks, the layout and ontologies of domain(in our case movies domain) get altered and thus affect the accuracy as shown in table 4. Also, the stochastic gradient descent technique used, due to noisy data takes significantly more time to gain convergence to the minima of the loss function. With an increase in noise, it also affects the prediction accuracy. The creation of neural networks is computationally expensive, and due to the noisy data issue of stochastic gradient descent, Paradarami approach takes considerably more time than our approach and Predictory approach, as seen in table 5.

Predictory [3] melds collaborative filtering techniques, content-based techniques, and the fuzzy expert system technique. Predictory's fuzzy expert system evaluated the significance of the movies with the help of parameters such as the number of ratings, the similarity measure between already rated movies, average movie rating, etc. Predictory paid more emphasis to the items than to the users. It failed to deliver more personalized recommendations. Therefore, compared to our model, it has a worse prediction accuracy. With the increase in noise due to the fuzzy expert system used in the model, the prediction accuracy abates, as seen in table 4. As it didn't have any Neural network creation or graph creation step, it has a better Mean Prediction Time than Paradarami RS and the other approaches.

Kamath et al.[7] used a smart news recommendation search engine that recognizes similar news articles available from numerous sources on a particular day; then, it applied a semantics-based enhanced bag-of-words technique to leverage the efficiency of search recommendation. Natural-language-processing(NLP) methods are used to identify articles dealing with similar news and serve on a user query with summarised news from various sources. For experimentation purposes, we use the Kamath model to search and recommend movies. The Kamath model's use of NLP helps it get better predictive accuracy than Paradarami and Predictory approach. But, the Kamath model suffers from the disadvantages of Natural language processing. It suffers from ambiguity and context problems while processing the reviews and other information extracted from IMDBPy. As a result, the prediction accuracy of Kamath is lower than Our approach and DNNRec. Moreover, the Kamath model performs Lexical Analysis, Syntactic Analysis, Semantic Analysis, Discourse Integration and Pragmatic Analysis on all the user and item data. Hence, it has a greater mean prediction time than mean prediction times of Our Approach, Predictory and Paradarami RS, as seen in table 5.

The DNNRec[6] utilizes embeddings for describing items and users to learn non-linear latent factors. Their approach uses embeddings and deep neural networks to find latent relations between users and items. Second, it retains traditional collaborative filtering techniques by learning latent factors and uses them to learn non-linear latent factors as embeddings. But, in doing so, it alters and, in the movies' case, does not maintain the domain structure, and hence

Table 5 Mean time taken (including Model Training time) for predictions at 30% noise level over 5 runs.

Approach	Mean Prediction Time (mins)
Paradarami [8]	1012
Predictory [3]	338
Kamath [7]	1116
DNNRec [6]	1278
Our Approach	**344**

it has worse efficiency than Our Approach. Our approach does not destroy the domain structure while creating user networks. Firstly, The Deep Neural Network used in DNNRec is very convoluted, and hence it requires a significant amount of time to generate. Secondly, the use of ReLU function in DNNRec leads to a positive mean shift in many layers, which in turn slows down learning. Therefore, due to the above 2 points, the mean prediction time of DNNRec is the highest among the five approaches.

5 Conclusion And Future Work

Previous works have shown that incorporating various approaches to form a hybrid model helps us enhance the accuracy of prediction. However the rating information is often noisy and sparse which may be misleading. Moreover, many CF based models are vulnerable to the cold start problem of a new user or an item and many deep learning based approaches are computationally expensive.

In this research, we presented a multi-layered hybrid approach in which we have incorporated item and user ontology in hybrid user-based and item-based CF model. Firstly, we applied SVD in traditional CF approach to resolve the sparsity issue of the rating matrix. Then we used both user network and item ontology to provide personalised recommendations even to the cold start and noisy users. In particular, we used DBSCAN algorithm, a density based clustering algorithm useful for large datasets. The major novelty in our approach is that we used a detailed and well defined movie ontology based on a domain-specific research to provide better results. Moreover, we considered user trust network analogous to item ontology and exploiting the explicit social links between users, we explored the preferences of connected users belonging to the same cluster in user based CF.

In the empirical evaluations, we can see that the multi-layered hybrid Re-CONet model performs better in terms of F score calculations than other neural network based, NLP-based and fuzzy expert syem based hybrid models. Since the primary aim of our recommender system is to provide personalised and relevant recommendations to users which are resistant to noise and sparsity issues, value of F1 score is relatively high and stable for different noise levels as compared to other models. It shows that Re-CONet is able to fulfill the desired aim. As shown in Table 4, our approach performed better under high noise levels and for 40% noise showed 18% F Score improvement over neural network

based model. The inter-cluster similarity graph finally approves that users having similar preferences belong to the same cluster and are given similar recommendations. The mean prediction time for our Model was also 66% less than Paradarami RS's time, 69% less than Kamath RS's time, 73% less than DNNRec's time and almost similar to Predictory's time .

In our future work, we propose to use clustering algorithms to incorporate new users and items in the clustering step itself. Since we are able to provide stable and relevant recommendations for a majorly sparse and noisy dataset of movie ratings, we assume that our model would be helpful in other domains such as tourism , literature, news, etc. In our future works we will create recommender system for e-learning domain which will help students learning from home get the right recommended topics and courses.

6 Declarations

6.1 Conflict of Interests and Competing Interests

- Employment: Deutsche Bank India, Ciena India Pvt Ltd, Snapdeal Pvt Ltd and Delhi Technological University
- Institution: Delhi Technological University

6.2 Financial Interests

- No competing financial interests in this research.

6.3 Funding

- This research was not externally funded and was self funded by the authors.

7 Data Availability Statements

The datasets are generated using publicly available datasets [34][35][36]. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

[1] Gomez-Uribe CA, Hunt N. The netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems (TMIS). 2015 Dec 28;6(4):1-9.

[2] Yin, Guisheng, et al. "Knowledge-aware recommendation model with dynamic co-attention and attribute regularize." Applied Intelligence (2021): 1-18.

[3] Walek B, Fojtik V. A hybrid recommender system for recommending relevant movies using an expert system. Expert Systems with Applications. 2020 Nov 15;158:113452.

[4] Moscato V, Picariello A, Sperli G. An emotional recommender system for music. IEEE Intelligent Systems. 2020 Sep 23.

[5] Mandal, Supriyo, and Abyayananda Maiti. "Deep collaborative filtering with social promoter score-based user-item interaction: a new perspective in recommendation." Applied Intelligence (2021): 1-26.

[6] Kiran, R., Pradeep Kumar, and Bharat Bhasker. "DNNRec: A novel deep learning based hybrid recommender system." Expert Systems with Applications 144 (2020): 113054.

[7] Kamath, S. Sowmya, and Monisha Kanakaraj. "Natural language processing-based e-news recommender system using information extraction and domain clustering." International Journal of Image Mining 1.1 (2015): 112-125.

[8] Paradarami TK, Bastian ND, Wightman JL. A hybrid recommender system using artificial neural networks. Expert Systems with Applications. 2017 Oct 15;83:300-13.

[9] Pujahari, Abinash, and Dilip Singh Sisodia. "Preference relation based collaborative filtering with graph aggregation for group recommender system." Applied Intelligence 51.2 (2021): 658-672.

[10] Luo X, Xia Y, Zhu Q. Incremental collaborative filtering recommender based on regularized matrix factorization. Knowledge-Based Systems. 2012 Mar 1;27:271-80.

[11] Luo X, Xia Y, Zhu Q. Applying the learning rate adaptation to the matrix factorization based Collaborative Filtering. Knowledge-Based Systems. 2013 Jan 1;37:154-64.

[12] Wang D, Liang Y, Xu D, Feng X, Guan R. A content-based recommender system for computer science publications. Knowledge-Based Systems. 2018 Oct 1;157:1-9.

[13] Nilashi M, Ibrahim O, Bagherifard K. A recommender system based on Collaborative Filtering using ontology and dimensionality reduction techniques. Expert Systems with Applications. 2018 Feb 1;92:507-20.

[14] Hsu, Chang-Ling. "A multi-valued and sequential-labeled decision tree method for recommending sequential patterns in cold-start situations." Applied Intelligence 51.1 (2021): 506-526.

[15] Middleton SE, De Roure D, Shadbolt NR. Ontology-based recommender systems. InHandbook on ontologies 2004 (pp. 477-498). Springer, Berlin, Heidelberg.

[16] Choudhury, Sasmita Subhadarsinee, Sachi Nandan Mohanty, and Alok Kumar Jagadev. "Multimodal trust based recommender system with machine learning approaches for movie recommendation." International Journal of Information Technology 13.2 (2021): 475-482.

[17] Wang Y, Yin G, Cai Z, Dong Y, Dong H. A trust-based probabilistic recommendation model for social networks. Journal of Network and Computer Applications. 2015 Sep 1;55:59-67.

[18] Kim Y, Shim K. TWILITE: A recommendation system for Twitter using a probabilistic model based on latent Dirichlet allocation. Information Systems. 2014 Jun 1;42:59-77.

[19] Chen, Lei, et al. "Improving the Prediction Quality in Memory-Based Collaborative Filtering Using Categorical Features." Electronics 10.2 (2021): 214.

[20] Di Noia T, Mirizzi R, Ostuni VC, Romito D. Exploiting the web of data in model-based recommender systems. InProceedings of the sixth ACM conference on Recommender systems 2012 Sep 9 (pp. 253-256).

[21] Sadaei HJ, Enayatifar R, Lee MH, Mahmud M. A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting. Applied Soft Computing. 2016 Mar 1;40:132-49.

[22] Gruber TR. A translation approach to portable ontology specifications. Knowledge acquisition. 1993 Jun 2;5(2):199-221.

[23] Daramola O, Adigun MO, Ayo CK. Building an ontology-based framework for tourism recommendation services. Building an Ontology-Based Framework for Tourism Recommendation Services. 2009:135-47.

[24] Hawalah A, Fasli M. Utilizing contextual ontological user profiles for personalized recommendations. Expert Systems with Applications. 2014 Aug 1;41(10):4777-97.

[25] Abdul-Rahman A, Hailes S. Supporting trust in virtual communities. InProceedings of the 33rd annual Hawaii international conference on system sciences 2000 Jan 7 (pp. 9-pp). IEEE.

[26] Li W, Qi J, Yu Z, Li D. A social recommendation method based on trust propagation and singular value decomposition. Journal of Intelligent & Fuzzy Systems. 2017 Jan 1;32(1):807-16.

[27] Wu J, Chiclana F, Fujita H, Herrera-Viedma E. A visual interaction consensus model for social network group decision making with trust propagation. Knowledge-Based Systems. 2017 Apr 15;122:39-50.

[28] Chen CC, Wan YH, Chung MC, Sun YC. An effective recommendation method for cold start new users using trust and distrust networks. Information Sciences. 2013 Mar 1;224:19-36.

[29] Gao P, Miao H, Baras JS, Golbeck J. Star: Semiring trust inference for trust-aware social recommenders. InProceedings of the 10th ACM Conference on Recommender Systems 2016 Sep 7 (pp. 301-308).

[30] Gohari FS, Haghighi H, Aliee FS. A semantic-enhanced trust based recommender system using ant colony optimization. Applied Intelligence. 2017 Mar 1;46(2):328-64.

[31] Jøsang A, Pope S. Semantic constraints for trust transitivity. InProceedings of the 2nd Asia-Pacific conference on Conceptual modelling-Volume 43 2005 Jan 1 (pp. 59-68).

[32] Balabanović M, Shoham Y. Fab: content-based, Collaborative recommendation. Communications of the ACM. 1997 Mar 1;40(3):66-72.

[33] Barragáns-Martínez AB, Costa-Montenegro E, Burguillo JC, et al. A hybrid content-based and item-based Collaborative Filtering approach to recommend TV programs enhanced with singular value decomposition. Information Sciences. 2010 Nov 15;180(22):4290-311.

[34] R. Zafarani and H. Liu, (2009). Social Computing Data Repository at ASU. Tempe, AZ: Arizona State University, School of Computing, Informatics and Decision Systems Engineering, http://socialcomputing.asu.edu.

[35] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4, Article 19 (December 2015), 19 pages. DOI=http://dx.doi.org/10.1145/2827872

[36] Butler, Martin, and Stefan Robila. "Interface for querying and data mining for the IMDb dataset." 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT). IEEE, 2016.

[37] Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering. 2007 Jan 1;60(1):208-21.

[38] Kristensen TG. Transforming Tanimoto queries on real valued vectors to range queries in Euclidian space. Journal of mathematical chemistry. 2010

Aug 1;48(2):287-9.

[39] Ceylan U, Birturk A. Content-boosted Collaborative Filtering using Semantic Similarity Measure. InWEBIST 2011 May 6 (pp. 366-371).

[40] Anowar, Farzana, Samira Sadaoui, and Bassant Selim. "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)." Computer Science Review 40 (2021): 100378.

View publication stats

arXiv:2308.07354v1 [gr-qc] 14 Aug 2023

A New Cosmological Model: Exploring the Evolution of the Universe and Unveiling Super-Accelerated Expansion

Himanshu Chaudhary,[1,2,3,*] Dhruv Arora,[2,†] Ujjal Debnath,[4,‡] G.Mustafa,[5,6,§] and S. K. Maurya[7,¶]

[1]*Department of Applied Mathematics, Delhi Technological University, Delhi-110042, India,*
[2]*Pacif Institute of Cosmology and Selfology (PICS), Sagara, Sambalpur 768224, Odisha, India*
[3]*Department of Mathematics, Shyamlal College, University of Delhi, Delhi-110032, India,*
[4]*Department of Mathematics, Indian Institute of Engineering Science and Technology, Shibpur, Howrah-711 103, India,*
[5]*Department of Physics, Zhejiang Normal University, Jinhua 321004, People's Republic of China,*
[6]*New Uzbekistan University, Mustaqillik ave. 54, 100007 Tashkent, Uzbekistan,*
[7]*Department of Mathematical and Physical Sciences, College of Arts and Sciences, University of Nizwa, Nizwa, Sultanate of Oman*

In this paper, we present a cosmological model designed to study the evolution of the universe based on a new parametrization of the deceleration parameter. The model considers a spatially flat, homogeneous, and isotropic Friedmann-Lemaître-Robertson-Walker (FLRW) universe filled with radiation, dark matter (DM), and dark energy (DE). We derive the Friedmann equations and the energy conservation equation for the universe, accounting for separate conservation equations for radiation, DM, and DE. Our proposed deceleration parameter is given by a formula involving constants H_0, Ω_{r0}, Ω_{m0}, q_2, q_1, q_0, α and β. which we subsequently fit to observational data. To assess the model's viability, we compare it with a diverse range of observational data, including cosmic chronometers, type Ia supernovae, baryon acoustic oscillations, and cosmic microwave background measurements. Employing the chi-square statistic and a Markov Chain Monte Carlo (MCMC) method, we estimate the best-fit values for the free parameters and investigate the constraints imposed by observational data on the model. Our results indicate that our cosmological model provides an excellent fit to the observed data and exhibits a remarkable agreement with the standard ΛCDM paradigm at higher redshifts. However, the most intriguing discovery lies in the model's prediction of a super-accelerated expansion in the distant future, in contrast to the de Sitter phase predicted by ΛCDM. This implies the presence of dark energy driving the universe's accelerated expansion. Additionally, the model displays unique behaviors in the jerk and snap parameters, indicating novel dynamics beyond the traditional cosmological model. These findings suggest that our proposed cosmological model offers a compelling alternative to the ΛCDM paradigm, shedding new light on the nature of dark energy and the future fate of the cosmos.

CONTENTS

I. Introduction 2

II. Fundamental Equations of the FLRW Model 3

III. Unveiling the Parameterized Deceleration: A Cosmic Odyssey 4
A. A New Deceleration Odyssey 4

IV. Data Analysis 5
A. Methodology : Markov Chain Monte Carlo Analysis 5
B. Data Discription 5
1. Cosmic Chronometers (CC) 5
2. type Ia supernovae (SNIa) 5
3. Baryon Acoustic Oscillations (BAO) 6
4. Cosmic Microwave Background 7

V. Observational, and theoretical comparisons with CC and SNIa Measurements 7
A. Contrasting with CC Measurements 7
B. Contrasting with type Ia supernovae (SNIa). 8
C. Relative differences between Model and ΛCDM. 9

VI. Cosmographic Parameters 9
A. The deceleration parameter 10
B. The jerk parameter 10
C. The snap parameter 11

VII. Statefinder Diagnostic 11

VIII. Om diagnostic. 12

IX. Information Criteria 12

X. Results 13

XI. Discussions and Conclusions 14

References 14

* himanshuch1729@gmail.com
† arora09dhruv@gmail.com
‡ ujjaldebnath@gmail.com
§ gmustafa3828@gmail.com
¶ sunil@unizwa.edu.om (corresponding author)

I. INTRODUCTION

The acceleration of cosmic expansion is seen to be one of the most important findings in modern cosmology [1]. The cosmos appears to be expanding with repulsive force and not slowing down under regular gravity. This unidentified factor, known as dark energy (DE), is responsible for current cosmic acceleration. DE is a mysterious mechanism of energy that is thought to pervade all of space and tends to burst the extension of the cosmos in physical cosmology and astronomy [2]. However, the nature of DE is yet uncertain, necessitating additional investigation. In the classic cosmological cold dark matter (CDM) model, the total mass energy of the universe consists of 4.9 % ordinary matter, 26.8 % DM, and 68.3 % dark energy. Also, in astronomy, dark matter (DM) is an undiscovered kind of substance that appears only as a gravitational interaction but does neither emit or absorb light [3]. The nature of DM is yet unclear, although its presence has been demonstrated by astronomical measurements [4, 5].

Various approaches have been suggested to explain DE, including quintessence [6], quintom [7], Chaplygin gas with its modified model [8], K-essence [9], new agegraphic DE [10], holographic DE model [11], pilgrim DE model [12], and Tsallis holographic DE (THDE) [13]. The simplest of them is the cosmological constant hypothesis, which is consistent with data [14]. Despite its success, it suffers from some problems such as the coincidence problem, the fine tuning problem and the age problem [15–17]. The equation of state (EoS) parameter ω defines the relationship between energy density and pressure in the cosmic framework. This is a dimensionless parameter that describes the cosmos' phases [18]. A parametrized version of ω is assumed for departing behavior of DE. A Taylor series expansion in the redshift, scale factor or any other parametrization of may be used [19].

The main idea in this approach is to consider a specific evolution scenario instead of considering any DE model apriori and then determine the nature of the exotic component that is triggering cosmic acceleration. It is known as the model-independent approach, which depends on estimating model parameters from existing observational datasets [20–25]. However, number of studies have tried to find out the history of the universe without relying on any particular cosmological model. Such techniques are frequently referred to as cosmography or cosmokinetic models, but we shall simply refer to them as kinematic models [26, 27]. This name stems from the fact that the entire study of the Universe's expansion (or its kinematics) is characterised solely by the Hubble expansion rate $H = \frac{\dot{a}}{a}$, the deceleration parameter $q = -\frac{a\ddot{a}}{\dot{a}^2}$ and the jerk parameter $j = -\frac{\dddot{a}\,a^2}{\dot{a}^3}$, where a is the scale factor in the Friedmann-Lemaître-Robertson-Walker (FLRW) metric. The deceleration parameter allows us to investigate the transition from a decelerated to an accelerated phase, whereas the jerk parameter allows us to investigate deviations from the cosmic concordance model without being limited to a certain model. In terms of the deceleration parameter, various studies have attempted to determine the redshift at which the universe transitions to an accelerated phase. [28–30]. In modern cosmology, A model-independent determination of the current deceleration parameter q_0 and the deceleration-acceleration transition redshift are two of the most important things.

It is required to utilise some parametrization for the deceleration parameter in order to investigate it in a cosmological-model independent context. This system has both benefits and drawbacks. One benefit is that it does not depend on the amount of matter and energy in the universe. One problem with this method is that it is unable to clarify the reason for accelerating expansion. Furthermore, the current deceleration parameter's value may be affected by the anticipated form of $q(z)$. Recently, Mamon et al. [31] studied a special form of deceleration parameter and obtained the best-fit values using χ^2 minimization technique with available observational data. They also analyzed the evolution of the jerk parameter for the considered parametrized model. Gadbali, Mandal, and Sahoo [32] have explored a specific parametrization of deceleration parameter in the context of $f(Q)$ gravity theory and constrained the model parameter by using Bayesian analysis with observational data. Much recently, sveral theoretical models have been developed to analyze the entire evolutionary history of the Universe through parametrization of $q(z)$ as a function of scale factor $(a(t))$ or time (t) or redshift in [33–72].

Motivated by the above developments, in this paper we present a parametrization of deceleration parameter with five unknown parameters. This work mainly focuses on constraining the model parameters using various observational datasets. In order to find the constraints on the cosmic parameters, we make use of 31 points of the Cosmic chronometers dataset, 1071 points of the supernova type Ia, 17 points of BAO and 162 points of the GRB dataset and 24 observations of the compact radio quasar obtained from a VLBI all-sky survey of 613 milliarcsecond ultra-compact radio sources at GHz. We have also made use of the CMB dataset. We adopt MCMC analysis to fit the model parameters on the data and perform data analysis by χ^2 minimization technique. Finally, for each model, we did a graphical analysis of cosmographic parameters such as deceleration, jerk, and snap. In addition to this we have discussed and analysed the statefinder and Om diagnostics for our parametrization.

The paper is structured as follows: section(II) introduces the basic equations of FLRW universe. Section(III) we discuss the Taylor series expansion of

deceleration parameter in terms of cosmic reshift, and introduce a new five parameter parametrization of DP. Thereafter, we find the Hubble function corresponding to the parametrization. In section(IV), the data description and analysis methodologies using MCMC techniques are presented to constrain the model parameters and find the best fit values. Section(V) compares the model's predictions with observational data. Section (VI) gives a detailed description about the kinematic cosmographic parameters such as the deceleration, jerk and snap parameters. Section(VII) and (VIII) discuss about the statefinder and Om diagnostics and present the evolution history of dark energy on $s-r$ and $q-r$ planes. Section(IX) discusses about the information criteria. Section(X) and Section(XI) discusses the results and conclusions respectively.

II. FUNDAMENTAL EQUATIONS OF THE FLRW MODEL

In this section, we delve into the foundational equations of the spatially flat, homogeneous, and isotropic FLRW (Friedmann-Lemaître-Robertson-Walker) universe. This model serves as a cornerstone in our exploration of the cosmos. We embark on our journey by embracing the line element that characterizes the very fabric of space and time:

$$ds^2 = -dt^2 + a^2(t)\left[dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right], \quad (1)$$

Here, $a(t)$ unfurls itself as the scale factor, a pivotal entity shaping the evolution of our cosmic theater. Within this cosmic arena, the energy-momentum tensor of the fluid orchestrates the intricate dance of matter and energy. It assumes the form:

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}, \quad (2)$$

At its core, ρ and p personify the energy density and pressure density of this cosmic fluid. The 4-velocity $u^\mu = \frac{dx^\mu}{ds}$ weaves the narrative of motion through the cosmos, steadfastly adhering to the profound relationship $u^\mu u_\mu = -1$. The FLRW Universe beckons us to uncover its secrets encoded in the Friedmann equations, sacred inscriptions within Einstein's gravitational framework.

Friedmann Equation 1:

$$H^2 = \frac{8\pi G}{3}\,\rho, \quad (3)$$

Friedmann Equation 2:

$$\dot{H} = -4\pi G(\rho + p), \quad (4)$$

Here, $H = \frac{\dot{a}}{a}$ emerges as the herald of cosmic expansion, and the temporal derivative denoted by the overhead dot gracefully embodies the cosmic passage of time, t. In the realm of these equations, we discern the harmonious interplay between energy densities, pressures, and the cosmic fabric itself, unveiling the essence of our FLRW cosmos. Delving into the intricate tapestry of the cosmos, we find its very essence woven with fluid matter, manifesting as a symphony of energy density ρ and pressure p. This cosmic ensemble adheres to a sacred edict, the energy conservation equation, resonating as:

$$\dot{\rho} + 3H(\rho + p) = 0, \quad (5)$$

Here, H is the celestial conductor, the Hubble parameter orchestrating the cosmic rhythm. Our odyssey into the cosmic composition unveils a triad of cosmic constituents: radiation, dark matter (DM), and enigmatic dark energy (DE). Their presence shapes the grand narrative of our universe, with densities and pressures woven into the cosmic fabric: $\rho = \rho_r + \rho_m + \rho_d$ and $p = p_r + p_m + p_d$. Peering into their individual realms, we invoke distinct conservation equations:

Radiation:

$$\dot{\rho}r + 3H(\rho r + p_r) = 0, \quad (6)$$

Dark Matter (DM):

$$\dot{\rho}m + 3H(\rho m + p_m) = 0, \quad (7)$$

Dark Energy (DE):

$$\dot{\rho}d + 3H(\rho d + p_d) = 0, \quad (8)$$

The radiant dance of radiation is characterized by $p_r = \frac{1}{3}\rho_r$, yielding the mesmerizing revelation $\rho_r = \rho_{r0}a^{-4}$, where a gracefully denotes the cosmic scale factor. Meanwhile, the enigmatic dark matter, a realm of negligible pressure ($p_m = 0$), weaves its destiny as $\rho_m = \rho_{m0}a^{-3}$. In this symphony of cosmic evolution, these equations breathe life into the dynamic interplay of cosmic components, echoing through the vast cosmic expansion. In our cosmic odyssey, the deceleration parameter, a compass guiding the cosmic voyage, unfurls before us:

$$q = -1 - \frac{\dot{H}}{H^2} \quad (9)$$

Venturing further, we unravel the narrative of the enigmatic dark energy. This enigma unfolds as:

$$q_d = -1 - \frac{\dot{H}_d}{H_d^2} \quad (10)$$

Here, H_d takes on the mantle of the cosmic maestro, conducting the expansion of the dark energy domain. A

cosmic duet of equations, (3) and (4), resonate harmoniously, revealing cosmic truths:

$$H_d^2 = \frac{8\pi G}{3}\ \rho_d \tag{11}$$

$$\dot{H}d = -4\pi G(\rho d + p_d) \tag{12}$$

Stepping onto the cosmic stage with these equations, the ethereal dance of fluid energy density takes center stage:

$$\rho_d = \rho_{d0}\ e^{\int \frac{2(1+q_d)}{1+z} dz} \tag{13}$$

A symphony of symbols and cosmic energies converge, where ρ_{d0} embodies the present density parameter, and z portrays the redshift parameter, narrated as $1+z=\frac{1}{a}$ (with $a_0 = 1$ in the present moment). In the cosmic chronicle, celestial constants emerge as cosmic constellations. Casting our gaze upon the cosmic canvas, we define the cosmic tapestry:

$\Omega_{r0} = \frac{8\pi G\rho_{r0}}{3H_0^2}$, $\Omega_{m0} = \frac{8\pi G\rho_{m0}}{3H_0^2}$, $\Omega_{d0} = \frac{8\pi G\rho_{d0}}{3H_0^2}$. From this cosmic calculus, the Hubble parameter script is penned:

$$H^2(z) = H_0^2\left[\Omega_{r0}(1+z)^4 + \Omega_{m0}(1+z)^3 + \Omega_{d0}\ e^{\int \frac{2(1+q_d)}{1+z} dz}\right] \tag{14}$$

In this cosmic sonnet, Ω_{d0} emerges as the sentinel of cosmic balance, whispering the cosmic tale of unity: $\Omega_{d0} = 1 - \Omega_{r0} - \Omega_{m0}$.

III. UNVEILING THE PARAMETERIZED DECELERATION: A COSMIC ODYSSEY

In our pursuit of deciphering the cosmos, we turn our gaze toward the elusive parameterized deceleration. Its simplest embodiment takes form as follows:

$$q(z) = q_0 + q_1\mathcal{X}(z) \tag{15}$$

Within this cosmic equation, q_0 and q_1 emerge as steadfast cosmic companions, while $\mathcal{X}(z)$, a function of redshift z, guides us through the cosmic terrain. Various cosmic riddles beckon, each demanding a distinct form for $\mathcal{X}(z)$. Yet, previous attempts often left cosmic enigmas unsolved, hinting at the need for a parametrization that truly captures the essence of cosmic evolution.responding Hubble parameter in terms of redshift z.

A. A New Deceleration Odyssey

In our cosmic odyssey, we unveil a new paradigm, a symphony of cosmic harmony:

$$q_d(z) = q_0 + \frac{\alpha + (1+z)^\beta}{q_1 + q_2(1+z)^\beta} \tag{16}$$

Within this cosmic symphony, q_0, q_1, q_2, α and β take on the role of cosmic constants, conducting the cosmic dance. This newfound paradigm ripples through the cosmic tapestry, influencing the very fabric of cosmic energy density:

$$\rho_d = \rho_{d0}\ (1+z)^{3(1+q_0+\frac{\alpha}{q_1})}[q_1 + q_2(1+z)^\beta]^{\frac{3(q_1-\alpha q_2)}{\beta q_1 q_2}} \tag{17}$$

This cosmic crescendo echoes through the cosmic chronicle, reshaping the Hubble parameter:

$$H^2(z) = H_0^2\left[(1+z)^4\,\Omega_{r0} + (1+z)^3\,\Omega_{m0} + (1-\Omega_{r0}-\Omega_{m0})\ (1+z)^{3(1+q_0+\frac{\alpha}{q_1})}[q_1 + q_2(1+z)^\beta]^{\frac{3(q_1-\alpha q_2)}{\beta q_1 q_2}}\right] \tag{18}$$

In this cosmic symphony, Ω_{r0} and Ω_{m0} compose the cosmic orchestra, weaving the cosmic melody that resonates across the cosmos. As we embark on this cosmic voyage, our new model of deceleration parameter charts a course toward understanding the enigmatic cosmic fabric.

IV. DATA ANALYSIS

In this section, we will conduct a thorough comparison between our proposed model's predictions and observational data. Our goal is to explore the constraints imposed on the model by utilizing two distinct observational datasets: the Cosmic Chronometers (CC) dataset, type Ia supernovae (SNIa) dataset, Gamma Ray Bursts (GRBs), Quasar (Q), Baryon Acoustic Oscillation (BAO) and Cosmic Microwave Background (CMB) measurements. By comparing the model's predictions with these cosmological data, we can determine the best-fit and mean values for the two free parameters of our model, denoted as Ω_{r0}, Ω_{m0}, q_2, q_1, q_0, α and β. It is crucial to incorporate the present-day value of the Hubble function H_0 into our analysis. To constrain the parameters of our cosmological model, we adopt a rigorous and widely-used approach based on Bayesian statistics. This technique involves the use of likelihood functions and the application of the MCMC method. The Bayesian framework provides a robust framework for parameter estimation by quantifying the probability of obtaining a certain set of model parameters given the observational data.

A. Methodology : Markov Chain Monte Carlo Analysis

In this section, We present our methodology to determine optimal values of the seven free parameters Ω_{r0}, Ω_{m0}, q_2, q_1, q_0, α, and β with present-day value of the Hubble function H_0, using MCMC. This advanced approach explores parameter space, inferring parameter values from observed data while quantifying uncertainties. Gaussian errors are assumed for each dataset. The likelihood, encapsulating data agreement, is expressed as:

$$\mathcal{L}(\theta) \propto \exp\left(-\frac{1}{2}\sum_i \frac{(x_{\mathrm{obs},i} - x_{\mathrm{th},i}(\theta))^2}{\sigma_i^2}\right), \tag{19}$$

Here, $\theta = (\Omega_{r0}, \Omega_{m0}, q_2, q_1, q_0, \alpha, \beta, H_0)$ denotes model parameters, $x_{\mathrm{obs},i}$ is observed data, $x_{\mathrm{th},i}(\theta)$ is model prediction, and σ_i is measurement standard deviation. The posterior distribution of parameters is governed by Bayes' theorem:

$$P(\theta|\mathrm{data}) \propto \mathcal{L}(\theta) \times \mathrm{Prior}(\theta). \tag{20}$$

We initiate by defining parameter space and employing suitable prior distributions, reflecting initial parameter beliefs. Likelihood, gauging model-data agreement, is computed considering data deviations. Bayesian inference combines priors with likelihood, yielding the posterior parameter distribution. MCMC algorithms, such as Metropolis Hastings [73] or Hamiltonian Monte Carlo [74], generate parameter samples from the posterior. These samples collectively traverse parameter space, converging towards higher likelihood regions. Convergence is ensured by monitoring diagnostics like Gelman-Rubin statistic [75] and trace plots. A burn-in phase is discarded, stabilizing chains and mitigating initial point influence. Best-fit parameter values are calculated from converged MCMC samples, with associated uncertainties as sample spreads. The New Model's performance is evaluated against diverse datasets, including Cosmic Chronometers (CC), Type Ia Supernovae (SNIa), Gamma Ray Bursts (GRBs), Quasars (Q), Baryon Acoustic Oscillations (BAO), and Cosmic Microwave Background (CMB). Model selection criteria, like Bayesian Information Criterion (BIC) [76] or Akaike Information Criterion (AIC) [77], aid in choosing the optimal model considering data fit and complexity. This methodology robustly explores parameter space, extracting the New Model's optimal parameters while accounting for uncertainties and observational data.

B. Data Discription

1. *Cosmic Chronometers (CC)*

The Hubble parameter, $H(z)$, can be determined by measuring the quantity dt, which is derived from a spectroscopic survey and represents the differential age technique. In this study, we utilize a collection of 31 data points, known as cosmic chronometers, to serve as constraints for our cosmological model. These data points provide valuable information about the expansion rate of the Universe at different redshifts [78, 79]. To assess the agreement between the theoretical predictions, denoted as $H_{\mathrm{th}}(z_i)$, and the observed values, denoted as $H_{\mathrm{obs}}(z_i)$, we employ the chi-square function:

$$\chi^2_{\mathrm{CC}} = \sum_{i=1}^{31} \frac{[H_{\mathrm{th}}(z_i) - H_{\mathrm{obs}}(z_i)]^2}{\sigma^2_{H(z_i)}},$$

where $\sigma_{H(z_i)}$ represents the standard error associated with the observed value of H. By minimizing the discrepancy between the theoretical predictions and the observed data using the chi-square function, we can determine the mean values of the model parameters. This makes it possible for us to find a quantitative estimate of the agreement between our cosmological model and the cosmic chronometer data.

2. *type Ia supernovae (SNIa)*

Over the years, Numerous supernova data has been established [80–84]. Recently, they took a fresh Pantheon sample, known as the Pantheon+ sample [85]. It consists of 1701 SNIa data points covering a redshift range from

$0.001 < z < 2.3$.. SNIa observations have played a key role in uncovering the fact regarding accelerating expansion of the universe. These observations widely used to study the nature of the component driving this expansion. SNIa are incredibly luminous astrophysical objects and are considered as standard candles, which means their intrinsic brightness can be used to measure relative distances. The Pantheon+ dataset provides valuable information for studying the nature of the accelerating universe. The chi-square statistic is commonly used to compare observational data with theoretical models. In the case of the Pantheon+ dataset, the chi-square values are calculated using the following equation:

$$\chi^2_{\rm Pantheon+} = \vec{D}^T \cdot \mathbf{C}^{-1}_{\rm Pantheon+} \cdot \vec{D} \tag{21}$$

Here, $\vec{D}$ represents the difference between the observed apparent magnitudes m_{Bi} of SNIa and the expected magnitudes given by the cosmological model. M represents the absolute magnitude of SNIa, and $\mu_{\rm model}$ is the corresponding distance modulus predicted by the assumed cosmological model. The term $\mathbf{C}_{\rm Pantheon+}$ denotes the covariance matrix provided with the Pantheon+ data, which includes both statistical and systematic uncertainties. The distance modulus is a measure of the distance to an object, defined as:

$$\mu_{\rm model}(z_i) = 5\log_{10}\left(\frac{D_L(z_i)}{(H_0/c)\ {\rm Mpc}}\right) + 25 \tag{22}$$

Here, $D_L(z)$ represents the luminosity distance, which is calculated for a flat homogeneous and isotropic FLRW universe as:

$$D_L(z) = (1+z)H_0 \int_0^z \frac{dz'}{H(z')} \tag{23}$$

The Pantheon+ dataset differs from the previous Pantheon sample as it breaks the degeneracy between the absolute magnitude M and the Hubble constant H_0. This is achieved by rewriting the vector $\vec{D}$ in terms of the distance moduli of SNIa in the Cepheid hosts. The distance moduli in the Cepheid hosts, denoted as $\mu_i^{\rm Ceph}$, are measured independently using Cepheid calibrators. This allows for the independent constraint of the absolute magnitude M. The modified vector $\vec{D}'$ is defined as:

$$\vec{D'}_i = \begin{cases} m_{Bi} - M - \mu_i^{\rm Ceph} & \text{if } i \text{ is in Cepheid hosts} \\ m_{Bi} - M - \mu_{\rm model}(z_i) & \text{otherwise} \end{cases} \tag{24}$$

With this modification, the chi-square equation for the Pantheon+ dataset can be rewritten as:

$$\chi^2_{\rm SN} = \vec{D'}^T \cdot \mathbf{C}^{-1}_{\rm Pantheon+} \cdot \vec{D'} \tag{25}$$

This revised formulation allows for improved constraints on the absolute magnitude M and the cosmological parameters.

We've also expanded our investigation to include a subset of 162 Gamma Ray Bursts (GRBs) [86], spanning a redshift range of $1.44 < z < 8.1$. In this context, we define the χ^2 function as:

$$\chi^2_{\rm GRB}(\phi^\nu_{\rm g}) = \mu_{\rm g}\mathbf{C}^{-1}_{\rm g,cov}\mu^T_{\rm g}, \tag{26}$$

Here, $\mu_{\rm g}$ denotes the vector encapsulating the differences between the observed and theoretical distance moduli for each individual GRB. Similarly, for our examination of 24 compact radio quasar observations [87], spanning redshifts in the range of $0.46 \leq z \leq 2.76$, we establish the χ^2 function as:

$$\chi^2_{\rm Q}(\phi^\nu_{\rm q}) = \mu_{\rm q}\mathbf{C}^{-1}_{\rm q,cov}\mu^T_{\rm q}, \tag{27}$$

In this context, $\mu_{\rm q}$ represents the vector capturing the disparities between the observed and theoretical distance moduli for each quasar.

3. *Baryon Acoustic Oscillations (BAO)*

To investigate the intricate phenomena of Baryon Acoustic Oscillations (BAO) in our study, we engage a comprehensive dataset encompassing 333 distinct measurements [88–99]. However, in light of potential uncertainties attributed to data correlations, we judiciously curate a more focused dataset comprising 17 meticulously selected BAO measurements for our analytical endeavors. This deliberate curation aims to mitigate errors and enhance the precision of our findings. A pivotal metric derived from BAO investigations along the transverse direction is the parameter $D_H(z)/r_d$, where $D_H(z)$ signifies the comoving angular diameter distance. This quantity is intimately linked to the following expressive relation [100, 101]:

$$D_M = \frac{c}{H_0} S_k\left(\int_0^z \frac{dz'}{E\,(z')}\right), \tag{28}$$

where $S_k(x)$ is defined as:

$$S_k(x) = \begin{cases} \frac{1}{\sqrt{\Omega_k}}\sinh\left(\sqrt{\Omega_k}x\right) & \text{if} \quad \Omega_k > 0 \\ x & \text{if} \quad \Omega_k = 0 \\ \frac{1}{\sqrt{-\Omega_k}}\sin\left(\sqrt{-\Omega_k}x\right) & \text{if} \quad \Omega_k < 0. \end{cases} \tag{29}$$

Furthermore, our exploration encompasses the incorporation of the angular diameter distance D_A, given by $D_A = D_M/(1+z)$, and the parameter $D_V(z)/r_d$. The latter amalgamates the BAO peak coordinates with r_d,

which characterizes the sound horizon during the drag epoch. Moreover, we gain valuable insights from "line-of-sight" or "radial" observations derived directly from the Hubble parameter, as expressed by:

$$D_V(z) \equiv \left[zD_H(z)D_M^2(z)\right]^{1/3}. \tag{30}$$

By studying these BAO measurements, we gain insights into the cosmological properties and evolution of the universe, while minimizing potential errors and considering relevant distance measures and observational parameters.

The total χ^2 function is given by the sum of the individual contributions:

4. *Cosmic Microwave Background*

In this study, distant measurements of the Cosmic Microwave Background (CMB) were analyzed [102]. These measurements provide valuable information about the CMB power spectrum in two aspects: the acoustic scale l_A characterizes the variation in peak spacing of the CMB temperature power spectrum in the transverse direction, and the "shift parameter" R influences the peak heights along the line-of-sight direction. The acoustic scale is defined as:

$$l_A = (1+z_d)\frac{\pi D_A(z)}{r_s}, \tag{31}$$

where z_d is the redshift at the drag epoch, $D_A(z)$ is the comoving angular diameter distance, and r_s is an independent parameter. The shift parameter is given by:

$$R(z) = \frac{\sqrt{\Omega_m}H_0}{c}(1+z_d)D_A(z), \tag{32}$$

here density parameter is denoted by Ω_m, while H_0 describes the Hubble constant and c denotes speed of light. The study [102] reports the following observables: $R_z = 1.7502 \pm 0.0046$, $l_A = 301.471 \pm 0.09$, and $n_s = 0.9649 \pm 0.0043$. The parameter r_s has an associated covariance matrix, which is detailed in Table I of [102]. These observables represent inflationary features and the expansion rate of the CMB epoch.

In addition to the CMB data, other data from the late Universe were also considered in the analysis. The combined results, including Cosmic Chronometers (CC), Type Ia supernovae (SNIa), Baryon Acoustic Oscillations (BAO), and the CMB, were used to test the model. The results of the analysis showed good agreement between the model and the data. The contour plots illustrating the combined results of CC, SNIa, BAO, and CMB are presented in Figure 1. Furthermore, the best-fit values along with their corresponding error bars are tabulated in Table I.

$$\chi^2_{\rm tot} = \chi^2_{CC} + \chi^2_{\rm SNIa} + \chi^2_{\rm GRB} + \chi^2_{\rm Q} + \chi^2_{BAO} + \chi^2_{CMB} \tag{33}$$

MCMC Results			
Model	Priors	Parameters	Best fit Value
ΛCDM Model	[50., 100.]	H_0	$69.854848^{+1.259100}_{-1.259100}$
	[0., 1.]	$\Omega_{\rm m0}$	$0.268654^{+0.012822}_{-0.012822}$
Model 1	[50., 100.]	H_0	$67.58035^{+2.358976}_{-1.141624}$
	[0., 1.]	Ω_{m0}	$0.21204^{+0.075725}_{-0.050090}$
	[0., 1.]	Ω_{r0}	$0.01045^{+0.005334}_{-0.004427}$
	[0., 1.]	q_2	$0.55564^{+0.012995}_{-0.009805}$
	[0., 1.]	q_1	$0.61298^{+0.057102}_{-0.031386}$
	[−3., −2.]	q_0	$-2.40901^{+0.877460}_{-0.640061}$
	[0., 1.]	α	$0.608808^{+0.432281}_{-0.221784}$
	[3., 4.]	β	$3.595096^{+0.594959}_{-0.320916}$

TABLE I. Summary of the MCMC results using CC + SNIa + Q + GRB + BAO + CMB dataset

V. OBSERVATIONAL, AND THEORETICAL COMPARISONS WITH CC AND SNIA MEASUREMENTS

After determining the values of the free parameters in our cosmological model, it is crucial to compare the model's predictions with observational data and the well-established ΛCDM paradigm. This step allows us to assess the model's viability and its ability to explain observed phenomena in the universe. By contrasting the model predictions with observational data, we can evaluate the agreement and identify any discrepancies or deviations from the standard ΛCDM framework. This analysis plays a crucial role in validating or refining the proposed cosmological model and advancing our understanding of the underlying physical processes driving the evolution of the universe.

A. Contrasting with CC Measurements

We assess the compatibility of our model with observational data by comparing it to the 31 measurements of Cosmic Chronometers datasets, represented by orange dots accompanied by error bars in the purple line shown in Figure 2. For reference, we also include the well-established ΛCDM model in black line with $\Omega_{\rm m0} = 0.3$ and $\Omega_\Lambda = 0.7$. The comparative analysis provides essential insights into how well our model performs in describing the observed Hubble data. The figure shows a close alignment between our model's predictions and the orange data points, indicating that our model successfully captures the features and trends present in the dataset. This alignment demonstrates that our model

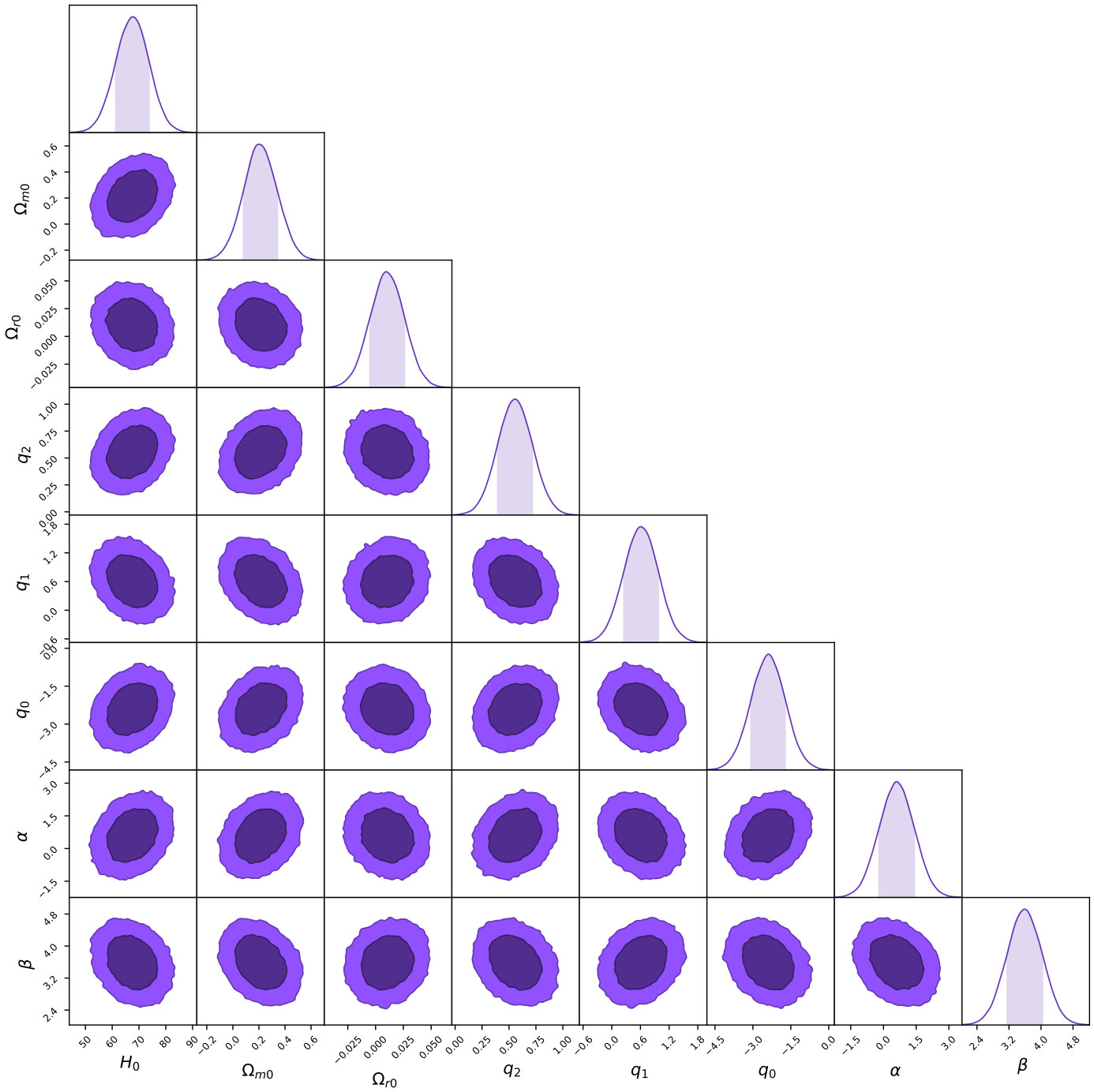


FIG. 1. This figure corresponds to 1σ and 2σ confidence contours obtained from $H(z)$ + SNIa + Q + GRB + BAO + CMB dataset obtained for the emergent model.

can reproduce the expansion history of the universe as inferred from the Hubble data. The agreement between our model and the observed cosmic evolution supports its validity and credibility.

B. Contrasting with type Ia supernovae (SNIa).

We compared the distance modulus function, denoted as $\mu(z)$, of our model with 1048 observational points from type Ia supernovae (SNIa) observations, shown in orange dots. The accompanying error bars represent measurement uncertainties. For reference, we included

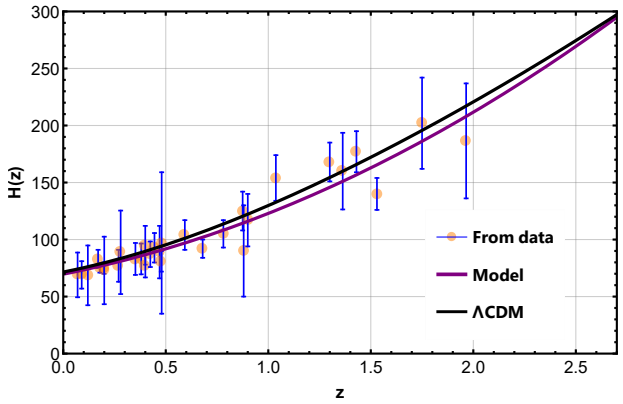


FIG. 2. Comparative analysis of our model (Purple line) with 31 Cosmic Chronometers datasets (orange dots) and ΛCDM model (black line)

the ΛCDM model as a black line. The distance modulus function $\mu(z)$ is a crucial quantity in cosmology, describing the apparent brightness of distant objects, such as supernovae, and enabling us to infer their distances. By comparing our model's predictions with the observed SNIa data, we can assess the model's ability to reproduce the observed behavior of the universe. Figure 3 visually demonstrates the close agreement between our model and the SNIa measurements. The fact that our model closely matches the dataset suggests that it provides an excellent fit for the observed supernova data. This agreement supports the viability of our model in describing the cosmological expansion and provides compelling evidence for its ability to explain the behavior of the universe as captured by the SNIa observations.

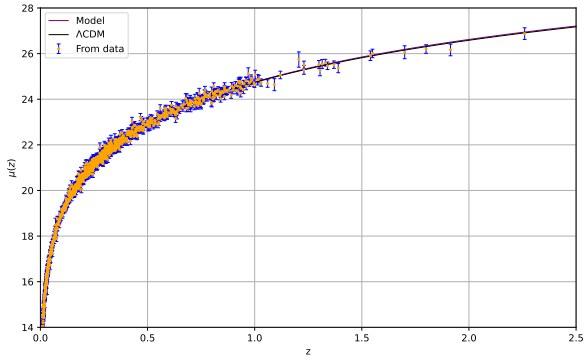


FIG. 3. Comparative analysis of our model (Purple line) with 1701 SNIa measurements (orange dots) and ΛCDM model (black line).

C. Relative differences between Model and ΛCDM.

The relative distinctions between our model and the ΛCDM model, we present Figure 4. The plot illustrates the behaviors of both models for redshifts $z < 1$, where they exhibit comparable dynamics. However, as the redshift surpasses unity, slight discrepancies emerge between the two models. These differences are particularly noticeable in the higher redshift range. The deviations between our model and the ΛCDM model gradually diminish beyond a critical redshift of approximately $z \approx 0.5$. This convergence at lower redshifts indicates that both models tend to align in their predictions as we explore earlier cosmic epochs. The close resemblance between our model and the ΛCDM model for $z < 1$ suggests that both models adequately capture the observed cosmic dynamics within this range. However, the observed discrepancies for higher redshifts indicate the presence of unique features or alternative mechanisms in our model, potentially influencing the universe's evolution at those epochs.

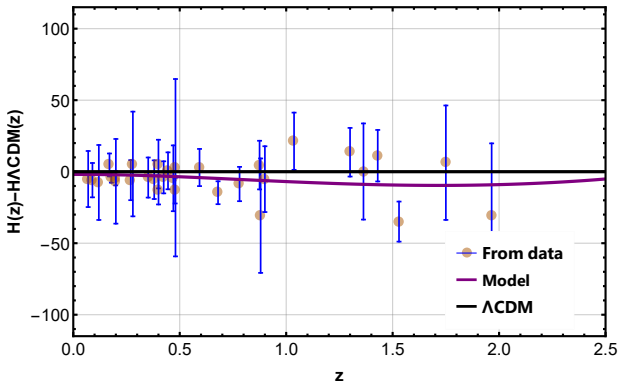


FIG. 4. Comparative analysis of our model (Purple line) with 31 Cosmic Chronometers datasets (orange dots) and ΛCDM model (black line)

VI. COSMOGRAPHIC PARAMETERS

Cosmographic parameters are mathematical quantities used to describe the behavior of the universe and its expansion. These parameters, such as the deceleration parameter, jerk parameter, and snap parameter, are derived from the Taylor series expansion of the scale factor and its derivatives.They give important insights into the dynamics and history of the cosmos, allowing us to investigate the acceleration or deceleration of its expansion as well as higher-order properties [19]. Researchers can acquire a better grasp of the underlying physics and test alternative hypotheses of dark energy by analysing and comparing these cosmographic features [103].

A. The deceleration parameter

In cosmology, the deceleration parameter is a dimensionless quantity used to determine the pace of expansion of the universe. It informs us about the accelerated or decelerated stages of cosmic expansion. It is defined mathematically as [104].

$$q = -\frac{a\ddot{a}}{\dot{a}^2}. \tag{34}$$

The DP influences the behavior of the Hubble parameter, which characterizes the rate at which the universe is expanding [105]. Depending on the sign of the deceleration parameter, the Hubble parameter can either increase or decrease. Positive values of q indicate an accelerated expansion of the universe, while negative values correspond to a decelerated expansion. Understanding and estimating the deceleration parameter is crucial for exploring the range of possible values denoted by q_0 through observational analysis. One way to probe the deceleration parameter is by studying the apparent brightness and redshift of supernovae in distant galaxies [106]. By examining these observational data, researchers can gain insights into the behavior of the deceleration parameter and its impact on the universe's expansion. Although estimating the deceleration parameter may seem challenging, recent findings strongly support models that suggest an accelerating universe [107]. These new outcomes have greatly favored the notion of an expanding universe with an accelerated rate of expansion. In cosmology, there is often a desire to determine a precise value for q_0, representing the deceleration parameter at the present epoch. However, achieving such precision requires rigorous observational analysis and careful consideration of various astrophysical phenomena [108]. Overall, the deceleration parameter provides valuable insights into the dynamics of cosmic expansion. By examining its behavior and exploring its implications through observational studies, scientists can deepen their understanding of the universe's evolution.

B. The jerk parameter

The dimensionless jerk parameter provides a generalization of the conventional cosmological parameters, such as the scale factor $a(t)$ and the deceleration parameter q. It arises from the fourth term in a Taylor series expansion of the scale factor around a reference time t_0. The expansion is given by:

$$\frac{a(t)}{a_0} = 1 + (t-t_0)\,H_0 - \frac{(t-t_0)^2}{2}\,H_0^2\,q_0 + \frac{(t-t_0)^3}{6}\,H_0^3\,j_0 + O\left[(t-t_0)^4\right], \tag{35}$$

where subscript 0 denotes the present values of the parameters. This expansion captures the behavior of the scale factor around t_0 up to the third derivative. Mathematically, the jerk parameter j is defined as the third derivative of the scale factor with respect to cosmic time, normalized by the ratio of the first derivative of the scale factor to the scale factor itself:

$$j = \frac{1}{a}\frac{d^3a}{d\tau^3}\left(\frac{1}{a}\frac{da}{d\tau}\right)^{-3} = q(2q+1) + (1+z)\frac{dq}{dz}. \tag{36}$$

The jerk parameter plays a significant role in distinguishing various proposals for dark energy (DE) and their implications for the dynamics of the universe. It helps identify favorable candidates for the physical interpretation of cosmic dynamics amidst the wide range of DE possibilities. By examining specific values of the jerk parameter, one can establish correspondences between DE proposals and standard models of the universe. Models involving a cosmic jerk provide a more pronounced demonstration of transitions between different eras of accelerated expansion. Notably, the flat ΛCDM model corresponds to a jerk parameter value of $j = 1$.

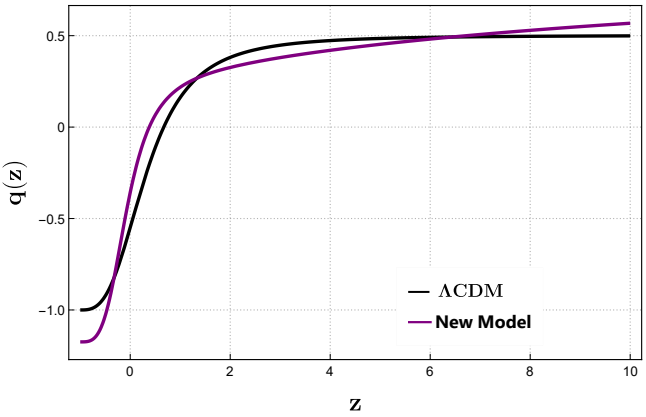


FIG. 5. This figure provides visual perception between respective and ΛCDM Model of deceleration Parameter.

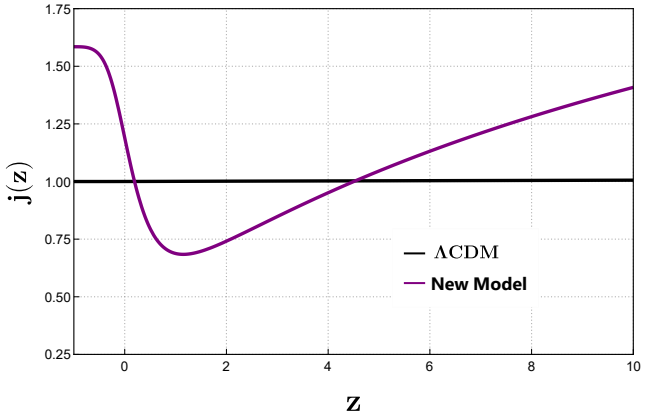


FIG. 6. This figure provides visual perception between respective and ΛCDM Model of jerk Parameter.

C. The snap parameter

The snap parameter, also known as jounce, is a dimensionless quantity that provides further insights into the behavior of the expansion factor. It is derived from the fifth term in the Taylor series expansion of the scale factor around the reference value a_0. The expansion is given by:

$$\frac{a(t)}{a_0} = 1 + H_0(t-t_0) - \frac{1}{2}q_0H_0^2(t-t_0)^2 + \frac{1}{6}j_0H_0^3(t-t_0)^3 + \frac{1}{24}s_0H_0^4(t-t_0)^4 + O\left[(t-t_0)^5\right]. \tag{37}$$

Mathematically, the snap parameter s is defined as the fourth derivative of the scale factor with respect to cosmic time, normalized by the ratio of the first derivative of the scale factor to the scale factor itself:

$$s = \frac{1}{a}\frac{d^4a}{d\tau^4}\left(\frac{1}{a}\frac{da}{d\tau}\right)^{-4} = \frac{j-1}{3\left(q-\frac{1}{2}\right)}. \tag{38}$$

For the flat ΛCDM model, where $j = 1$, the snap parameter can be expressed as $s = -(2+3q)$. This relationship provides a clear connection between the snap parameter and the dynamics of the ΛCDM model. Furthermore, the divergence of $\frac{ds}{dq}$ from -3 indicates deviations in the evolutionary mechanism from the dynamics of the ΛCDM model. By examining the behavior of this derivative, we can gain insights into how the system deviates from the standard ΛCDM dynamics.

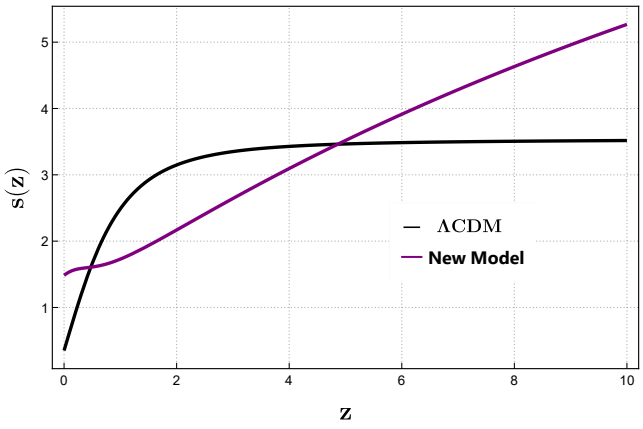


FIG. 7. This figure provides visual perception between respective and ΛCDM Model of snap Parameter.

VII. STATEFINDER DIAGNOSTIC

The statefinder diagnostic pair $\{r, s\}$ is a widely used tool in the study of dark energy (DE) models, offering insights into their nature based on higher-order derivatives of the scale factor [109]. These dimensionless parameters provide a model-independent way to analyze the cosmic properties of DE. The statefinder pair is computed using the following expressions:

$$r = \frac{\dddot{a}}{aH^3}, \quad s = \frac{r-1}{3\left(q-\frac{1}{2}\right)}, \tag{39}$$

Here, r represents the ratio of the third derivative of the scale factor to the cube of the Hubble parameter, and s is a linear combination of r and the deceleration parameter q. Specific values of r and s have well-established interpretations within standard DE models. For example, $\{r, s\} = \{1, 0\}$ corresponds to the ΛCDM model, while $\{r, s\} = \{1, 1\}$ corresponds to the standard cold dark matter (SCDM) model in a Friedmann-Lemaître-Robertson-Walker (FLRW) universe. The value range of $(-\infty, \infty)$ indicates the possibility of an Einstein static universe. In the $r - s$ plane, different regions correspond to distinct types of DE models. For instance, $s > 0$ signifies quintessence-like models, whereas $s < 0$ indicates phantom-like models. Deviations from the standard value of $\{r, s\} = \{1, 0\}$ can indicate an evolutionary process from phantom-like to quintessence-like behavior. Moreover, specific combinations of the deceleration parameter q and the statefinder parameter r are associated with well-known models. The pair $\{q, r\} = \{-1, 1\}$ is linked to the ΛCDM model, while $\{q, r\} = \{0.5, 1\}$ corresponds to the SCDM model.

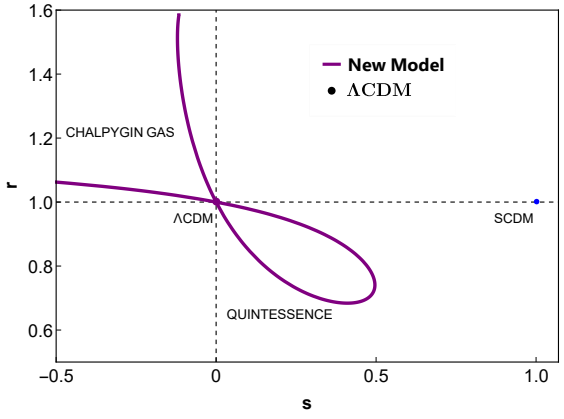


FIG. 8. Behavior of $\{r, s\}$ plane

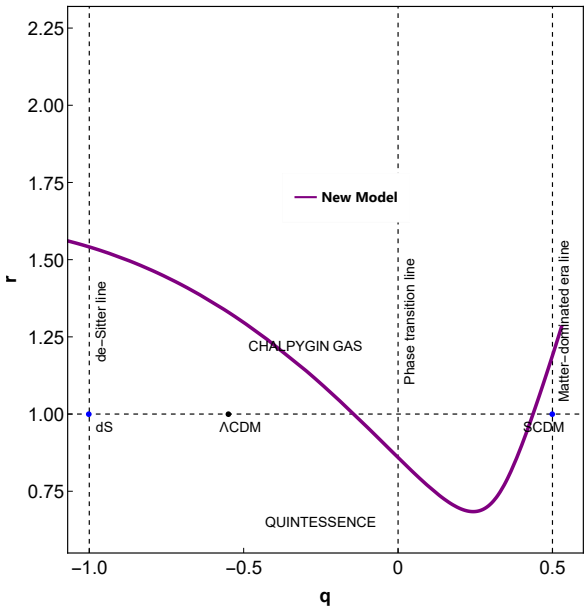


FIG. 9. Behavior of $\{r, q\}$ plane

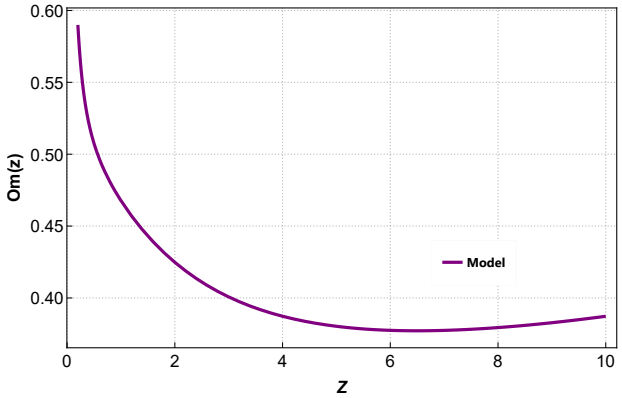


FIG. 10. Behavior of $Om(z)$ Profile

VIII. OM DIAGNOSTIC.

The Om diagnostic is a geometrical formalism used to test the validity of the ΛCDM model in cosmology [110]. It provides a useful tool for differentiating between various dark energy (DE) models and the cosmological constant. By analyzing the slope variation of the Om parameter with redshift, one can distinguish quintessence and phantom models based on positive and negative slopes, respectively. In the case of a constant slope, it corresponds to the cosmological constant. In a flat universe, the Om diagnostic is defined as the ratio of the squared Hubble parameter at a given redshift to its present value, normalized by the expansion factor.

$$Om(z) = \frac{\left(\frac{H(z)}{H_0}\right)^2 - 1}{(1+z)^3 - 1}. \tag{40}$$

Unlike the statefinder diagnosis, the Om diagnostic only requires the first-order temporal derivative. It has also been applied in the study of Galileon models. Overall, the Om diagnostic provides a valuable tool for exploring and understanding different DE scenarios in cosmology.

IX. INFORMATION CRITERIA

Model selection criteria such as the corrected Akaike information criterion (AIC) play a golden role in evaluating the sustainability of the models [77, 111–114]. The AIC_c takes into account the maximum likelihood function, the number of data, and the number of free parameters of models. For larger datasets, the AIC_c reduces to AIC. The AIC is used to figure out the goodness-of-fit of different models. The model having the lowest AIC is the most preferred by observations. The difference, $\Delta AIC = AIC_{preferred} - AIC_{model}$, is more meaningful that the AIC itself. A range of $(0, 2)$ for ΔAIC refers to models strongly supported, values in the range of $(4, 7)$ are less supported, while values beyond 10 refer to models without any support. Therefore, the use of AIC is a very useful tool for classifying a set of models based on their ability to fit data. In addition to the AIC, the Bayesian information criterion (BIC) [76, 115, 116] also contributes to model selection. Like the AIC, the BIC considers the trade-off between goodness-of-fit and model complexity. However, the BIC incorporates a stronger penalty for models with more parameters. This promotes the selection of simpler models, helping to guard against overfitting. In the context of ΔBIC, similar principles apply as with ΔAIC. A lower ΔBIC indicates stronger support for a particular model, and the magnitude of the difference is informative

Model	χ^2_{min}	χ^2_{red}	AIC	ΔAIC	BIC	ΔBIC
ΛCDM	1742.41	0.9876	1748.41	0	1763.94	0
Model	1733.01	0.9864	1749.01	0.69084	1790.41	26.4774

TABLE II. Summary of $\chi^2_{\rm tot}{}^{min}$, $\chi^2_{\rm red}$, AIC, ΔAIC, BIC, ΔBIC

X. RESULTS

a. deceleration parameter Figure 5 illustrates the changes in the deceleration parameter (DP) with respect to the cosmic redshift (z) in our new cosmological model. Remarkably, our model demonstrates an excellent agreement with the standard ΛCDM model at higher redshifts. However, the real excitement lies in the future phase of the universe's evolution, where our new model showcases a fascinating phenomenon - a super accelerated expansion ($q(-1) \approx -0.3$). This stands in stark contrast to the ΛCDM model, which predicts a de Sitter phase with $q(-1) = -1$. The discovery of this super-accelerated expansion has significant implications, as it suggests that the universe might experience unprecedented and rapid expansion in the distant future. Indeed, the observed super-accelerated expansion in the new cosmological model could potentially be attributed to dark energy.

b. jerk parameter Figure 6 shows the evolution of the jerk parameter with cosmic redshift (z) in the context of Cosmology. Our new model exhibits a remarkable departure from the well-established ΛCDM model. The jerk parameter initially experiences a steady decline during the early stages of the universe. However, this decline is eventually halted as the model intersects with the ΛCDM curve around $z \sim 4.25$. Subsequently, an intriguing reversal in the evolution of the jerk parameter occurs, reaching its minimum value at $z \sim 1$. After this critical point, our model aligns closely with the ΛCDM predictions, particularly near the present time, i.e., $z \sim 0$, where the universe continues to display accelerating expansion.

c. snap parameter We have examined the behavior of the snap parameter in comparison to the widely accepted ΛCDM model. Remarkably, our new model exhibits a significant and consistent departure from the ΛCDM predictions across all ranges of cosmic redshifts (z) as shown in Figure 7. The snap parameter experiences a steady and continuous decline, showcasing a unique pattern not seen in the standard model. Particularly noteworthy is the behavior of the snap parameter near the present time, at approximately $z \sim 0$. Our model predicts that the slope of the curve flattens out, leading to an intriguing prediction of $s(0) \sim 1.6$ for the snap parameter. This deviation from the ΛCDM model provides valuable insights into the dynamics of the universe and suggests alternative mechanisms that might govern cosmic expansion.

d. $\{r, s\}$ profile In Figure 8, we present the intriguing evolution of the $\{r, s\}$ parameter. Our model depicts an interesting journey through several dark energy domains, offering information on the universe's dynamics. Our model begins in the Chalpygin gas-type dark energy area, which is defined by $r > 1$ and $s0$. After this, the model crosses the fixed ΛCDM point $\{1, 0\}$ as the cosmic evolution advances, signalling a momentous change. Following that, it reaches the Quintessence type dark energy zone, with $\{r, s\}$ parameter values of $r < 1$ and $s > 0$. This phase suggests a transition away from the classic cosmological model and towards a different type of dark energy behaviour. Our model subsequently goes through another transition, returning to the Chalpygin gas-type dark energy area. This back-and-forth movement across the fixed ΛCDM point highlights the complexity and richness of the dynamics inherent in our proposed model.

e. $\{r, q\}$ profile Figure 9 illustrates the intriguing evolution of the $\{r, q\}$ parameter in our Cosmological model. At the beginning of the cosmic evolution, our model resides in the Chalpygin gas-dominated DE region, characterized by $r > 1$ and $q > 0$. This phase represents a unique state of the universe, influenced by specific properties of the Chalpygin gas. As cosmic time progresses, the model enters the Quintessence dominated DE region characterized by $r < 1$ and $q > 0$ after crossing the line associated with constant jerk value of $j = 1$. Following this, the model shows a phase transition from deceleration to acceleration and gradually enters the Chalpygin gas DE region again.

f. om Diagnostic In Figure 10, we present the interesting trend of the $Om(z)$ diagnostic with cosmic redshift z. The plot displays a consistent positive slope for all values of redshift, suggesting a Quintessence-type nature of dark energy. This finding indicates that the universe's expansion is driven by a dynamic and evolving form of dark energy. However, what makes the results even more fascinating is the brief interval of cosmic redshift, specifically $4.5 \leq z \leq 6.5$, where the slope of $Om(z)$ becomes nearly zero. This intriguing behavior hints at the presence of cosmological constant-type dark energy during this specific range of redshifts.

g. Information Criteria Based on the values presented in Table II, we can provide a comprehensive comparison between the New Model and the ΛCDM Model using various statistical measures, including $\chi^2_{\rm min}$, $\chi^2_{\rm red}$, AIC, ΔBIC, BIC, and ΔBIC. The $\chi^2_{\rm min}$ for the New Model is lower than that of the ΛCDM Model, indicating a better fit of the New Model to the data in terms of minimizing the total χ^2. The $\chi^2_{\rm red}$ values for both models are quite similar, suggesting that both models provide reasonable fits to the data when considering the number of degrees of freedom. The AIC value for the New Model is higher than that of

the ΛCDM Model, indicating that the ΛCDM Model has a better balance between goodness of fit and model complexity according to AIC. The calculated ΔAIC of 0.69084 suggests that the New Model is moderately supported compared to the ΛCDM Model. This value falls within the range of $(0, 2)$, indicating that the New Model has some level of support according to ΔAIC. The BIC value for the New Model is higher than that of the ΛCDM Model, reinforcing the idea that the ΛCDM Model is preferred in terms of model complexity and goodness of fit according to BIC. The ΔBIC value of 26.4774 suggests moderate support for the New Model over the ΛCDM Model, as it falls within the range of $(4, 7)$. This indicates that the New Model is less favored compared to the ΛCDM Model based on BIC. The evidence suggests that the ΛCDM Model has a better fit and is moderately supported over the New Model. The lower values of $\chi^2_{\min}$ and AIC for the ΛCDM Model indicate its improved goodness of fit and model selection. Additionally, the calculated ΔBIC value of 26.4774 indicates that the New Model is less favored compared to the ΛCDM Model, falling within the range of models that are less supported according to BIC.

XI. DISCUSSIONS AND CONCLUSIONS

The authors of the paper presented a cosmological model to study the evolution of the Universe based on a new parametrization of the deceleration parameter. The model considered a spatially flat, homogeneous, and isotropic Friedmann-Lemaître-Robertson-Walker (FLRW) universe filled with radiation, dark matter (DM), and dark energy (DE). They derived the Friedmann equations and the energy conservation equation for the universe and considered separate conservation equations for radiation, DM, and DE. The new model proposed for the deceleration parameter was given as

$$q_d(z) = q_0 + \frac{\alpha + (1+z)^\beta}{q_1 + q_2(1+z)^\beta}$$

where q_0, q_1, q_2, α, and β are constants. The model was then tested against observational data from various sources, including cosmic chronometers, type Ia supernovae, baryon acoustic oscillations, and cosmic microwave background measurements. The comparison was done using the chi-square statistic, and a MCMC method was employed for parameter estimation and uncertainty estimation. The analysis showed that the model's predictions were in good agreement with the observational data. The best-fit values of the free parameters Ω_{r0}, Ω_{m0}, q_2, q_1, q_0, α, and β were obtained using the MCMC method. These values allowed the authors to understand the constraints imposed by the observational data on the model. We also compared our proposed cosmological model with observational data and the well-established ΛCDM paradigm. We found that our model's predictions aligned closely with the observed data from Cosmic Chronometers and type Ia supernovae (SNIa), indicating its ability to describe the universe's expansion history. Additionally, we identified slight discrepancies between our model and the ΛCDM model at higher redshifts, suggesting the presence of unique features in our model at those epochs. The cosmographic results of our new cosmological model demonstrate its remarkable ability to describe the observed universe and its expansion history. The model exhibits excellent agreement with the standard ΛCDM paradigm at higher redshifts. However, the real excitement lies in its prediction of a super-accelerated expansion in the distant future, in contrast to the de Sitter phase predicted by ΛCDM. This finding implies the presence of dark energy driving the accelerated expansion. The evolution of the jerk parameter in our model shows intriguing behavior, with a reversal in its decline around $z \sim 4.25$, leading to a minimum value around $z \sim 1$. This departure from the ΛCDM model highlights unique dynamics in our proposed model. Similarly, the snap parameter displays a continuous decline, deviating consistently from the ΛCDM predictions, particularly near the present time. The $\{r, s\}$ and $\{r, q\}$ profiles demonstrate fascinating journeys through different dark energy regions, indicating transitions between Chalpygin gas-type and Quintessence-type dark energy behaviors. These transitions further differentiate our model from the standard ΛCDM. The $Om(z)$ diagnostic suggests that our model is driven by Quintessence-type dark energy with a brief interval of nearly constant behavior hinting at cosmological constant-type dark energy at specific redshifts. Regarding the information criteria, our model has slightly less support than the ΛCDM model, but both models adequately fit the observed data. Our new cosmological model presents a compelling alternative to the ΛCDM paradigm, capturing unique aspects of cosmic expansion and offering valuable insights into the nature of dark energy.

[1] Adam G Riess, Alexei V Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M Garnavich, Ron L Gilliland, Craig J Hogan, Saurabh Jha, Robert P Kirshner, et al. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The astronomical journal*, 116(3):1009, 1998.

[2] P. J. E. Peebles and Bharat Ratra. The cosmological constant and dark energy. *Rev. Mod. Phys.*, 75:559–606, Apr 2003.

[3] Gianfranco Bertone and Dan Hooper. History of dark matter. *Reviews of Modern Physics*, 90(4):045002, 2018.

[4] David N Spergel and Paul J Steinhardt. Observational evidence for self-interacting cold dark matter. *Physical review letters*, 84(17):3760, 2000.

[5] Bin Wang, Elcio Abdalla, Fernando Atrio-Barandela, and Diego Pavon. Dark matter and dark energy interactions: theoretical challenges, cosmological implications and observational signatures. *Reports on Progress in Physics*, 79(9):096901, 2016.

[6] Paul J Steinhardt. A quintessential introduction to dark energy. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 361(1812):2497–2513, 2003.

[7] Hao Wei, Rong-Gen Cai, and Ding-Fang Zeng. Hessence: a new view of quintom dark energy. *Classical and Quantum Gravity*, 22(16):3189, 2005.

[8] Abha Dev, JS Alcaniz, and Deepak Jain. Cosmological consequences of a chaplygin gas dark energy. *Physical Review D*, 67(2):023515, 2003.

[9] Nilok Bose and AS Majumdar. K-essence model of inflation, dark matter, and dark energy. *Physical Review D*, 79(10):103517, 2009.

[10] Hao Wei and Rong-Gen Cai. A new model of agegraphic dark energy. *Physics Letters B*, 660(3):113–117, 2008.

[11] Shuang Wang, Yi Wang, and Miao Li. Holographic dark energy. *Physics reports*, 696:1–57, 2017.

[12] M Sharif and Abdul Jawad. Analysis of pilgrim dark energy models. *The European Physical Journal C*, 73:1–8, 2013.

[13] Emmanuel N Saridakis, Kazuharu Bamba, R Myrzakulov, and Fotios K Anagnostopoulos. Holographic dark energy through tsallis entropy. *Journal of Cosmology and Astroparticle Physics*, 2018(12):012, 2018.

[14] Luca Visinelli, Sunny Vagnozzi, and Ulf Danielsson. Revisiting a negative cosmological constant from low-redshift data. *Symmetry*, 11(8):1035, 2019.

[15] Nima Arkani-Hamed, Lawrence J Hall, Christopher Kolda, and Hitoshi Murayama. New perspective on cosmic coincidence problems. *Physical Review Letters*, 85(21):4434, 2000.

[16] Miguel Quartin, Mauricio O Calvao, Sergio E Joras, Ribamar RR Reis, and Ioav Waga. Dark interactions and cosmological fine-tuning. *Journal of Cosmology and Astroparticle Physics*, 2008(05):007, 2008.

[17] Jun Chen, Puxun Wu, Hongwei Yu, and Zhengxiang Li. Age problem in the creation cold dark matter cosmology model. *The European Physical Journal C*, 72:1–7, 2012.

[18] Robert R Caldwell, Rahul Dave, and Paul J Steinhardt. Cosmological imprint of an energy component with general equation of state. *Physical Review Letters*, 80(8):1582, 1998.

[19] Matt Visser. Jerk, snap and the cosmological equation of state. *Classical and Quantum Gravity*, 21(11):2603, 2004.

[20] Sergio Del Campo, Ivan Duran, Ramon Herrera, and Diego Pavón. Three thermodynamically based parametrizations of the deceleration parameter. *Physical Review D*, 86(8):083509, 2012.

[21] SD Katore and AY Shaikh. Anisotropic dark energy model with a special form of deceleration parameter. *Bulg. J. Phys*, 42:29–41, 2015.

[22] Adam G Riess, Louis-Gregory Strolger, John Tonry, Stefano Casertano, Henry C Ferguson, Bahram Mobasher, Peter Challis, Alexei V Filippenko, Saurabh Jha, Weidong Li, et al. Type ia supernova discoveries at z¿ 1 from the hubble space telescope: Evidence for past deceleration and constraints on dark energy evolution. *The Astrophysical Journal*, 607(2):665, 2004.

[23] Amine Bouali, Himanshu Chaudhary, Ujjal Debnath, Alok Sardar, and G Mustafa. Data analysis of three parameter models of deceleration parameter in frw universe. *arXiv preprint arXiv:2304.13137*, 2023.

[24] Amine Bouali, Himanshu Chaudhary, Ujjal Debnath, Tanusree Roy, and G Mustafa. Constraints on the parameterized deceleration parameter in frw universe. *arXiv preprint arXiv:2301.12107*, 2023.

[25] Amine Bouali, BK Shukla, Himanshu Chaudhary, Rishi Kumar Tiwari, Mahvish Samar, and G Mustafa. Cosmological tests of parametrization q= α- β h in f (q) flrw cosmology. *International Journal of Geometric Methods in Modern Physics*, page 2350152, 2023.

[26] Salvatore Capozziello, Rocco D'Agostino, and Orlando Luongo. Kinematic model-independent reconstruction of palatini f (r) cosmology. *General Relativity and Gravitation*, 51:1–19, 2019.

[27] Micol Benetti and Salvatore Capozziello. Connecting early and late epochs by f (z) cdm cosmography. *Journal of Cosmology and Astroparticle Physics*, 2019(12):008, 2019.

[28] Abdulla Al Mamon. Constraints on a generalized deceleration parameter from cosmic chronometers. *Modern Physics Letters A*, 33(10n11):1850056, 2018.

[29] Özgür Akarsu, Tekin Dereli, Suresh Kumar, and Lixin Xu. Probing kinematics and fate of the universe with linearly time-varying deceleration parameter. *The European Physical Journal Plus*, 129:1–14, 2014.

[30] Lixin Xu and Hongya Liu. Constraints to deceleration parameters by recent cosmic observations. *Modern Physics Letters A*, 23(23):1939–1948, 2008.

[31] Abdulla Al Mamon and Sudipta Das. A parametric reconstruction of the deceleration parameter. *The European Physical Journal C*, 77(7):495, 2017.

[32] Gaurav N Gadbail, Sanjay Mandal, and Pradyumn Kumar Sahoo. Parametrization of deceleration parameter in f (q) gravity. *Physics*, 4(4):1403–1412, 2022.

[33] Salvatore Capozziello, Rocco D'Agostino, and Orlando Luongo. Model-independent reconstruction of f (t) teleparallel cosmology. *General Relativity and Gravitation*, 49:1–21, 2017.

[34] Salvatore Capozziello, Rocco D'Agostino, and Orlando Luongo. Kinematic model-independent reconstruction of palatini f (r) cosmology. *General Relativity and Gravitation*, 51:1–19, 2019.

[35] Salvatore Capozziello, Rocco D'Agostino, and Orlando Luongo. Thermodynamic parametrization of dark energy. *Physics of the Dark Universe*, 36:101045, 2022.

[36] Salvatore Capozziello, Rocco D'Agostino, and Orlando Luongo. Cosmographic analysis with Chebyshev polynomials. *Mon. Not. Roy. Astron. Soc.*, 476(3):3924–3938, 2018.

[37] S. Capozziello, R. D'Agostino, and O. Luongo. High-redshift cosmography: auxiliary variables versus Padé polynomials. *Mon. Not. Roy. Astron. Soc.*, 494(2):2576–2590, 2020.

[38] Lorenzo Sebastiani, Sunny Vagnozzi, Ratbay Myrzakulov, et al. Mimetic gravity: a review of recent develop-

ments and applications to cosmology and astrophysics. *Advances in High Energy Physics*, 2017, 2017.

[39] Suhail Dhawan, Justin Alsing, and Sunny Vagnozzi. Non-parametric spatial curvature inference using late-universe cosmological probes. *Monthly Notices of the Royal Astronomical Society: Letters*, 506(1):L1–L5, 2021.

[40] Sunny Vagnozzi. Implications of the nanograv results for inflation. *Monthly Notices of the Royal Astronomical Society: Letters*, 502(1):L11–L15, 2021.

[41] Sunny Vagnozzi. Cosmological searches for the neutrino mass scale and mass ordering. *arXiv preprint arXiv:1907.08010*, 2019.

[42] Sunny Vagnozzi, Suhail Dhawan, Martina Gerbino, Katherine Freese, Ariel Goobar, and Olga Mena. Constraints on the sum of the neutrino masses in dynamical dark energy models with $w(z) \geq -1$ are tighter than those obtained in ΛCDM. *Phys. Rev. D*, 98(8):083501, 2018.

[43] Orlando Luongo. Cosmography with the Hubble parameter. *Mod. Phys. Lett. A*, 26:1459–1466, 2011.

[44] Orlando Luongo. Dark energy from a positive jerk parameter. *Mod. Phys. Lett. A*, 28:1350080, 2013.

[45] Orlando Luongo and Hernando Quevedo. A Unified Dark Energy Model from a Vanishing Speed of Sound with Emergent Cosmological Constant. *Int. J. Mod. Phys. D*, 23:1450012, 2014.

[46] Orlando Luongo, Giovanni Battista Pisani, and Antonio Troisi. Cosmological degeneracy versus cosmography: a cosmographic dark energy model. *Int. J. Mod. Phys. D*, 26(03):1750015, 2016.

[47] Orlando Luongo and Marco Muccino. Speeding up the universe using dust with pressure. *Phys. Rev. D*, 98(10):103520, 2018.

[48] Alejandro Aviles, Christine Gruber, Orlando Luongo, and Hernando Quevedo. Cosmography and constraints on the equation of state of the Universe in various parametrizations. *Phys. Rev. D*, 86:123516, 2012.

[49] Alejandro Aviles, Alessandro Bravetti, Salvatore Capozziello, and Orlando Luongo. Precision cosmology with Padé rational approximations: Theoretical predictions versus observational limits. *Phys. Rev. D*, 90(4):043531, 2014.

[50] Christine Gruber and Orlando Luongo. Cosmographic analysis of the equation of state of the universe through Padé approximations. *Phys. Rev. D*, 89(10):103506, 2014.

[51] Alejandro Aviles, Jaime Klapp, and Orlando Luongo. Toward unbiased estimations of the statefinder parameters. *Phys. Dark Univ.*, 17:25–37, 2017.

[52] Alejandro Aviles, Jaime Klapp, and Orlando Luongo. Toward unbiased estimations of the statefinder parameters. *Phys. Dark Univ.*, 17:25–37, 2017.

[53] Sergio del Campo, Ivan Duran, Ramon Herrera, and Diego Pavon. Three thermodynamically-based parameterizations of the deceleration parameter. *Phys. Rev. D*, 86:083509, 2012.

[54] JV Cunha and Jose Ademir Sales de Lima. Transition redshift: new kinematic constraints from supernovae. *Monthly Notices of the Royal Astronomical Society*, 390(1):210–217, 2008.

[55] J. V. Cunha. Kinematic Constraints to the Transition Redshift from SNe Ia Union Data. *Phys. Rev. D*, 79:047301, 2009.

[56] Adam G. Riess et al. Type Ia supernova discoveries at z $>$ 1 from the Hubble Space Telescope: Evidence for past deceleration and constraints on dark energy evolution. *Astrophys. J.*, 607:665–687, 2004.

[57] Lix-In Xu, Cheng-Wu Zhang, Bao-Rong Chang, and Hong-Ya Liu. Constraints to deceleration parameters by recent cosmic observations. *Mod. Phys. Lett. A*, 23:1939–1948, 2008.

[58] Lixin Xu and Jianbo Lu. Cosmic constraints on deceleration parameter with Sne Ia and CMB. *Mod. Phys. Lett. A*, 24:369–376, 2009.

[59] Remya Nair, Sanjay Jhingan, and Deepak Jain. Cosmokinetics: a joint analysis of standard candles, rulers and cosmic clocks. *Journal of Cosmology and Astroparticle Physics*, 2012(01):018, 2012.

[60] Ozgur Akarsu, Tekin Dereli, Suresh Kumar, and Lixin Xu. Probing kinematics and fate of the Universe with linearly time-varying deceleration parameter. *Eur. Phys. J. Plus*, 129:22, 2014.

[61] Beethoven Santos, Joel C Carvalho, and Jailson S Alcaniz. Current constraints on the epoch of cosmic acceleration. *Astroparticle Physics*, 35(1):17–20, 2011.

[62] Yun-Gui Gong and Anzhong Wang. Reconstruction of the deceleration parameter and the equation of state of dark energy. *Phys. Rev. D*, 75:043520, 2007.

[63] Michael S. Turner and Adam G. Riess. Do SNe Ia provide direct evidence for past deceleration of the universe? *Astrophys. J.*, 569:18, 2002.

[64] Abdulla Al Mamon and Sudipta Das. A divergence free parametrization of deceleration parameter for scalar field dark energy. *Int. J. Mod. Phys. D*, 25(03):1650032, 2016.

[65] Amine Bouali, Himanshu Chaudhary, Ujjal Debnath, Tanusree Roy, and G Mustafa. Constraints on the parameterized deceleration parameter in frw universe. *arXiv preprint arXiv:2301.12107*, 2023.

[66] Himanshu Chaudhary, Amine Bouali, Ujjal Debnath, Tanusree Roy, and Ghulam Mustafa. Constraints on the parameterized deceleration parameter in frw universe. *Physica Scripta*, 2023.

[67] Amine Bouali, Himanshu Chaudhary, Ujjal Debnath, Alok Sardar, and G Mustafa. Data analysis of three parameter models of deceleration parameter in frw universe. *arXiv preprint arXiv:2304.13137*, 2023.

[68] Amine Bouali, BK Shukla, Himanshu Chaudhary, Rishi Kumar Tiwari, Mahvish Samar, and G Mustafa. Cosmological tests of parametrization q= α- β h in f (q) flrw cosmology. *International Journal of Geometric Methods in Modern Physics*, page 2350152, 2023.

[69] Amine Bouali, Himanshu Chaudhary, Saadia Mumtaz, G Mustafa, and SK Maurya. Observational constraining study of new deceleration parameters in frw universe. *Fortschritte der Physik*, page 2300033, 2023.

[70] Amine Bouali, BK Shukla, Himanshu Chaudhary, Rishi Kumar Tiwari, and Marco San Martin. Cosmographic studies of q= α- β h parametrization in f (t) framework. *International Journal of Geometric Methods in Modern Physics*, 2023.

[71] Himanshu Chaudhary, Aditya Kaushik, and Ankita Kohli. Cosmological test of $\sigma\theta$ as function of scale factor in f (r, t) framework. *New Astronomy*, 103:102044, 2023.

[72] Amine Bouali, Himanshu Chaudhary, Amritansh Mehrotra, and SKJ Pacif. Model-independent study

for a quintessence model of dark energy: Analysis and observational constraints. *arXiv preprint arXiv:2304.02652*, 2023.

[73] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[74] Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*, 2017.

[75] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[76] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

[77] Scott I Vrieze. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228, 2012.

[78] Michele Moresco, Andrea Cimatti, Raul Jimenez, Lucia Pozzetti, Gianni Zamorani, Micoll Bolzonella, James Dunlop, Fabrice Lamareille, Marco Mignoli, Henry Pearce, et al. Improved constraints on the expansion rate of the universe up to z 1.1 from the spectroscopic evolution of cosmic chronometers. *Journal of Cosmology and Astroparticle Physics*, 2012(08):006, 2012.

[79] Michele Moresco, Lucia Pozzetti, Andrea Cimatti, Raul Jimenez, Claudia Maraston, Licia Verde, Daniel Thomas, Annalisa Citro, Rita Tojeiro, and David Wilkinson. A 6% measurement of the hubble parameter at z 0.45: direct evidence of the epoch of cosmic re-acceleration. *Journal of Cosmology and Astroparticle Physics*, 2016(05):014, 2016.

[80] Marek Kowalski, David Rubin, Greg Aldering, RJ Agostinho, A Amadon, R Amanullah, C Balland, K Barbary, G Blanc, PJ Challis, et al. Improved cosmological constraints from new, old, and combined supernova data sets. *The Astrophysical Journal*, 686(2):749, 2008.

[81] Rahman Amanullah, Chris Lidman, D Rubin, G Aldering, P Astier, K Barbary, MS Burns, A Conley, KS Dawson, SE Deustua, et al. Spectra and hubble space telescope light curves of six type ia supernovae at 0.511¡ z¡ 1.12 and the union2 compilation. *The Astrophysical Journal*, 716(1):712, 2010.

[82] N Suzuki, D Rubin, C Lidman, G Aldering, R Amanullah, K Barbary, LF Barrientos, J Botyanszki, M Brodwin, N Connolly, et al. The hubble space telescope cluster supernova survey. v. improving the dark-energy constraints above z¿ 1 and building an early-type-hosted supernova sample. *The Astrophysical Journal*, 746(1):85, 2012.

[83] MEA Betoule, R Kessler, J Guy, J Mosher, D Hardin, R Biswas, P Astier, P El-Hage, M Konig, S Kuhlmann, et al. Improved cosmological constraints from a joint analysis of the sdss-ii and snls supernova samples. *Astronomy & Astrophysics*, 568:A22, 2014.

[84] Daniel Moshe Scolnic, DO Jones, A Rest, YC Pan, R Chornock, RJ Foley, ME Huber, R Kessler, Gautham Narayan, AG Riess, et al. The complete light-curve sample of spectroscopically confirmed sne ia from pan-starrs1 and cosmological constraints from the combined pantheon sample. *The Astrophysical Journal*, 859(2):101, 2018.

[85] Dan Scolnic, Dillon Brout, Anthony Carr, Adam G Riess, Tamara M Davis, Arianna Dwomoh, David O Jones, Noor Ali, Pranav Charvu, Rebecca Chen, et al. The pantheon+ type ia supernova sample: the full dataset and light-curve release. *arXiv preprint arXiv:2112.03863*, 2021.

[86] Marek Demianski, Ester Piedipalumbo, Disha Sawant, and Lorenzo Amati. Cosmology with gamma-ray bursts-i. the hubble diagram through the calibrated ep, i–eiso correlation. *Astronomy & Astrophysics*, 598:A112, 2017.

[87] Carl Roberts, Keith Horne, Alistair O Hodson, and Alasdair Dorkenoo Leggat. Tests of \ λ cdm and conformal gravity using grb and quasars as standard candles out to z \ sim 8. *arXiv preprint arXiv:1711.10369*, 2017.

[88] Will J Percival, Beth A Reid, Daniel J Eisenstein, Neta A Bahcall, Tamas Budavari, Joshua A Frieman, Masataka Fukugita, James E Gunn, Zeljko Ivezi'c, Gillian R Knapp, et al. Baryon acoustic oscillations in the sloan digital sky survey data release 7 galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 401(4):2148–2168, 2010.

[89] Florian Beutler, Chris Blake, Matthew Colless, D Heath Jones, Lister Staveley-Smith, Lachlan Campbell, Quentin Parker, Will Saunders, and Fred Watson. The 6df galaxy survey: baryon acoustic oscillations and the local hubble constant. *Monthly Notices of the Royal Astronomical Society*, 416(4):3017–3032, 2011.

[90] Timothee Delubac, James Rich, Stephen Bailey, Andreu Font-Ribera, David Kirkby, J-M Le Goff, Matthew M Pieri, Anze Slosar, Éric Aubourg, Julian E Bautista, et al. Baryon acoustic oscillations in the lyα forest of boss quasars. *Astronomy & Astrophysics*, 552:A96, 2013.

[91] Lauren Anderson, Eric Aubourg, Stephen Bailey, Dmitry Bizyaev, Michael Blanton, Adam S Bolton, Jon Brinkmann, Joel R Brownstein, Angela Burden, Antonio J Cuesta, et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample. *Monthly Notices of the Royal Astronomical Society*, 427(4):3435–3467, 2012.

[92] Hee-Jong Seo, Shirley Ho, Martin White, Antonio J Cuesta, Ashley J Ross, Shun Saito, Beth Reid, Nikhil Padmanabhan, Will J Percival, Roland De Putter, et al. Acoustic scale from the angular power spectra of sdss-iii dr8 photometric luminous galaxies. *The Astrophysical Journal*, 761(1):13, 2012.

[93] Ashley J Ross, Lado Samushia, Cullan Howlett, Will J Percival, Angela Burden, and Marc Manera. The clustering of the sdss dr7 main galaxy sample–i. a 4 per cent distance measure at z= 0.15. *Monthly Notices of the Royal Astronomical Society*, 449(1):835–847, 2015.

[94] Rita Tojeiro, Ashley J Ross, Angela Burden, Lado Samushia, Marc Manera, Will J Percival, Florian Beutler, J Brinkmann, Joel R Brownstein, Antonio J Cuesta, et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: galaxy clustering measurements in the low-redshift sample of data release 11. *Monthly Notices of the Royal Astronomical Society*, 440(3):2222–2237, 2014.

[95] Julian E Bautista, Mariana Vargas-Magaña, Kyle S Dawson, Will J Percival, Jonathan Brinkmann, Joel Brownstein, Benjamin Camacho, Johan Comparat,

Hector Gil-Marín, Eva-Maria Mueller, et al. The sdss-iv extended baryon oscillation spectroscopic survey: baryon acoustic oscillations at redshift of 0.72 with the dr14 luminous red galaxy sample. *The Astrophysical Journal*, 863(1):110, 2018.

[96] E De Carvalho, A Bernui, GC Carvalho, CP Novaes, and HS Xavier. Angular baryon acoustic oscillation measure at z= 2.225 from the sdss quasar survey. *Journal of Cosmology and Astroparticle Physics*, 2018(04):064, 2018.

[97] Metin Ata, Falk Baumgarten, Julian Bautista, Florian Beutler, Dmitry Bizyaev, Michael R Blanton, Jonathan A Blazek, Adam S Bolton, Jonathan Brinkmann, Joel R Brownstein, et al. The clustering of the sdss-iv extended baryon oscillation spectroscopic survey dr14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2. *Monthly Notices of the Royal Astronomical Society*, 473(4):4773–4794, 2018.

[98] TMC Abbott, FB Abdalla, A Alarcon, S Allam, F Andrade-Oliveira, J Annis, S Avila, Mandakranta Banerji, N Banik, K Bechtol, et al. Dark energy survey year 1 results: Measurement of the baryon acoustic oscillation scale in the distribution of galaxies to redshift 1. *Monthly Notices of the Royal Astronomical Society*, 483(4):4866–4883, 2019.

[99] Z Molavi and A Khodam-Mohammadi. Observational tests of gauss-bonnet like dark energy model. *The European Physical Journal Plus*, 134(6):254, 2019.

[100] Natalie B Hogg, Matteo Martinelli, and Savvas Nesseris. Constraints on the distance duality relation with standard sirens. *Journal of Cosmology and Astroparticle Physics*, 2020(12):019, 2020.

[101] Matteo Martinelli, Carlos Jose Amaro Parente Martins, S Nesseris, Domenico Sapone, I Tutusaus, Anastasios Avgoustidis, Stefano Camera, Carmelita Carbone, S Casas, Stéphane Ilić, et al. Euclid: Forecast constraints on the cosmic distance duality relation with complementary external probes. *Astronomy & Astrophysics*, 644:A80, 2020.

[102] Rong-Gen Cai, Shi Pi, Shao-Jiang Wang, and Xing-Yu Yang. Pulsar timing array constraints on the induced gravitational waves. *Journal of Cosmology and Astroparticle Physics*, 2019(10):059, 2019.

[103] Kazuharu Bamba, Salvatore Capozziello, Shin'ichi Nojiri, and Sergei D Odintsov. Dark energy cosmology: the equivalent description via different theoretical models and cosmography tests. *Astrophysics and Space Science*, 342:155–228, 2012.

[104] Yu L Bolotin, VA Cherkaskiy, OA Lemets, DA Yerokhin, and LG Zazunov. Cosmology in terms of the deceleration parameter. part i. *arXiv preprint arXiv:1502.00811*, 2015.

[105] Salvatore Capozziello, Peter KS Dunsby, and Orlando Luongo. Model-independent reconstruction of cosmological accelerated–decelerated phase. *Monthly Notices of the Royal Astronomical Society*, 509(4):5399–5415, 2022.

[106] Moncy V John. Cosmographic evaluation of the deceleration parameter using type ia supernova data. *The Astrophysical Journal*, 614(1):1, 2004.

[107] Devaraja Mallesha Naik, NS Kavya, and V Venkatesha. Observational insights into the accelerating universe through reconstruction of the deceleration parameter. *Chinese Physics C*, 2023.

[108] David Camarena and Valerio Marra. Local determination of the hubble constant and the deceleration parameter. *Physical Review Research*, 2(1):013028, 2020.

[109] Varun Sahni, Tarun Deep Saini, Alexei A Starobinsky, and Ujjaini Alam. Statefinder—a new geometrical diagnostic of dark energy. *Journal of Experimental and Theoretical Physics Letters*, 77:201–206, 2003.

[110] Varun Sahni, Arman Shafieloo, and Alexei A Starobinsky. Two new diagnostics of dark energy. *Physical Review D*, 78(10):103502, 2008.

[111] MYJ Tan and Rahul Biswas. The reliability of the akaike information criterion method in cosmological model selection. *Monthly Notices of the Royal Astronomical Society*, 419(4):3292–3303, 2012.

[112] Mehdi Rezaei and Mohammad Malekjani. Comparison between different methods of model selection in cosmology. *The European Physical Journal Plus*, 136(2):219, 2021.

[113] Fabiola Arevalo, Antonella Cid, and Jorge Moya. Aic and bic for cosmological interacting scenarios. *The European Physical Journal C*, 77:1–13, 2017.

[114] Keith Tauscher, David Rapetti, and Jack O Burns. A new goodness-of-fit statistic and its application to 21-cm cosmology. *Journal of Cosmology and Astroparticle Physics*, 2018(12):015, 2018.

[115] Fotios K Anagnostopoulos, Spyros Basilakos, and Emmanuel N Saridakis. Bayesian analysis of f (t) gravity using f σ 8 data. *Physical Review D*, 100(8):083517, 2019.

[116] Fotios K Anagnostopoulos, Spyros Basilakos, and Emmanuel N Saridakis. Observational constraints on myrzakulov gravity. *Physical Review D*, 103(10):104013, 2021.

RSC Advances

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *RSC Adv.*, 2023, **13**, 24450

A novel cobalt(II) acetate complex bearing lutidine ligand: a promising electrocatalyst for oxygen evolution reaction†

Deepika Tanwar,‡[ab] Priya Jain,‡[c] Deepali Ahluwalia,[d] Athul Sudheendranath,[c] Sajesh P. Thomas,[c] Pravin P. Ingole *[c] and Umesh Kumar *[a]

Developing cost-effective electrocatalysts using earth-abundant metal as an alternative to expensive precious metal catalyst remains a key challenge for researchers. Several strategies are being researched/tested for making low-cost transition metal complexes with controlled electron-density and coordination flexibility around the metal center to enhance their catalytic activity. Herein, we report a novel lutidine coordinated cobalt(II) acetate complex $[(3,5\text{-lutidine})_2Co(OAc)_2(H_2O)_2]$ (**1**) as a promising electrocatalyst for oxygen evolution reaction (OER). Complex **1** was characterized by FT-IR, elemental analysis, and single crystal X-ray diffraction data. The structure optimization of complex **1** was also done using DFT calculation and the obtained geometrical parameters were found to be in good agreement with the parameters obtained from the solid state structure obtained through single crystal X-ray diffraction data. Further, the molecular electrostatic potential (MEP) maps analysis of complex **1** observed electron rich centers that were found to be in agreement with the solid-state structure. It was understood that the coordination of lutidine as a Lewis base and acetate moiety as a flexible ligand will provide more coordination flexibility around the metal center to facilitate the catalytic reaction. Further, the electron rich centers around metal center will also support the enhancement of their catalytic activity. Complex **1** shows impressive OER activity, even better than the state-of-the-art IrO_2 catalyst, in terms of turnover frequency (TOF: 0.05) and onset potential (1.50 V *vs.* RHE). The TOF for complex **1** is two and half times higher, while the onset potential is *ca.* 20 mV lower, than the benchmark IrO_2 catalyst studied under identical conditions.

Received 13th July 2023
Accepted 7th August 2023

DOI: 10.1039/d3ra04709a

rsc.li/rsc-advances

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Introduction

The development of cost-effective and clean energy is of considerable scientific interest due to the rapid increase in energy consumption and environmental crisis associated with fossil fuels.[1–3] Electrochemical water splitting is one of the promising strategies, which involves the hydrogen evolution reaction (HER) and oxygen evolution reaction (OER).[4] However, the commercialization of these devices is largely limited due to the high over-potential and sluggish kinetics of the OER, which requires highly active electrocatalysts.[5] Henceforth, in the last few decades, many research groups have been conducting their work in this area to develop efficient electrocatalysts for OER. Initially, precious metals (Pt, Ir and Ru) based catalysts were focused due to their robust nature and high activity.[6–10] In the recent past, the demand for cost-effective earth-abundant transition metal based electrocatalysts has accelerated to overcome the high cost and lower abundance of precious metals.[11–14] However, the OER activity of these transition metal-based catalysts has not been comparable to the Pt-group metals. Hence, the foremost challenge for chemists is to design and develop novel cost-effective, earth abundant and highly efficient catalysts as well as strategies to boost their activity for OER.

Cobalt-based catalysts have been significantly studied to facilitate the OER.[15,16] For instance, cobalt oxides,[17–20] hydroxides,[21,22] sulfides,[23–25] selenides,[26,27] phosphates/phosphides,[28,29] and nitrides[30,31] have been established as advanced electrocatalyst for OER. Beside these materials, several molecular cobalt complexes have also been found as efficient electrocatalyst for OER in recent times.[32–38] D. G. Nocera's group reported a β-octafluoro Co(III) xanthene hangman corrole as the effective OER catalyst under operating conditions at

[a]*Catalysis and Bioinorganic Research Lab, Department of Chemistry, Deshbandhu College, University of Delhi, New Delhi-110019, India. E-mail: ukumar@db.du.ac.in*
[b]*Department of Chemistry, University of Delhi, New Delhi-110007, India*
[c]*Department of Chemistry, Indian Institute of Technology, New Delhi-110016, India. E-mail: ppingole@chemistry.iitd.ac.in*
[d]*Department of Applied Chemistry, Delhi Technological University, New Delhi-110042, India*

† Electronic supplementary information (ESI) available: Crystallographic data for the complex **1** CCDC 2204239. For ESI and crystallographic data in CIF or other electronic format see DOI: https://doi.org/10.1039/d3ra04709a

‡ Deepika Tanwar and Priya Jain contributed equally to this work.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

modest overpotential.[32] J. T. Groves *et al.* reported a cationic cobalt–porphyrin complex as an efficient homogeneous electrocatalyst for water oxidation to study the role of buffer on the reactivity.[33] C. N. Verani *et al.* designed a novel phenolate rich cobalt(III) complex which acts as an excellent water oxidation catalyst at moderate overpotential of 0.5 V.[34] L. Sun *et al.* used cobalt porphyrin complexes as OER catalysts and found that the *in situ* generation of cobalt oxides thin film acts as a real catalyst.[35] Recently, M. Nath *et al.* reported a cobalt complex bearing selelno-based ligand as OER electrocatalyst with very low overpotential of 320 mV.[37] Zhen-Tao Yu *et al.* reported a cobalt(IV) diacetato complex bearing a substituted bipyridine dianionic ligand as highly active electrocatalyst for OER at an overpotential of only 360 mV at pH = 6.[38] In case of molecular cobalt catalyst, many research groups reported that the *in situ* formation of $Co(OH)_2$/cobalt oxide are the true catalyst.[19,35,39]

However, the nature of *in situ* generated catalyst depends upon the molecular nature of parent catalyst and hence the efficacy of the catalyst solely depends on role of parent complex. Overall, these representative examples of Co-complex based OER catalysts highlights the chemistry of ligands towards controlling the electron-density around the Co-metal center or an ability to partially oxidize the complex to make thin films of cobalt oxide. Hence, it was thought that the Lewis base ligand can control the electron-density around metal center to enhance their catalytic activity. Further, the coordination of Lewis base and acetate moiety will provide more coordination flexibility around metal center which supports the substrate coordination/removal during catalysis to facilitate the reaction cycle.[38,40] In this view, we report a novel cost-effective lutidine coordinated cobalt(II) acetate complex $[(3,5\text{-lutidine})_2Co(OAc)_2(H_2O)_2]$ (**1**) as promising electrocatalytic activator for OER. The obtained data after OER electrocatalytic analysis indicates that complex **1** is better performer than the state-of-the-art IrO_2 catalyst.

Experimental

Materials and methods

The details of materials and methods have been given in ESI.† The details of X-ray diffraction data collection and crystallographic information are also given in the ESI (see Table S1†).

Synthesis of complex 1

$Co(OAc)_2 \cdot 4H_2O$ (250 mg, 1 mmol) was dissolved in acetonitrile to get homogeneous solution, to which 3,5-lutidine (214 mg, 2 mmol in 5 mL acetonitrile) was added dropwise and stirred at room temperature for 3 h. The resulting solution was reduced using rotary evaporator and then kept at room temperature for crystallization. Crystals suitable for single crystal X-ray diffraction data of $[(3,5\text{-lutidine})_2Co(OAc)_2(H_2O)_2]$ (**1**) were obtained after five days. Yield: 94% (0.402 g, 0.940 mmol). FT-IR (KBr, ν in cm^{-1}): 3425 (br) for $\nu(H_2O)$; 1572 (s) for $\nu_{asym}(OCO)$; and 1419 (s) for $\nu_{sym}(OCO)$. Anal. calcd (%) for $C_{18}H_{28}CoN_2O_6$: C, 50.59; H, 6.60; N, 6.56. Found C, 51.02; H, 6.57; N, 6.71.

Electrochemical analysis

The electrochemical analysis for oxygen evolution reaction was performed in a three-electrode electrochemical setup, using Glassy carbon (GC) electrode with geometric area of 0.07 cm^2 as working electrode, Hg/HgO (1 M NaOH) ($E^0 = 0.118$ V *vs.* RHE) as reference electrode and graphite rod as counter electrode in 1 M KOH (pH = 14). The GC electrode was modified with catalyst ink prepared by dispersing the catalyst in 3 : 1 100 μL isopropanol and water mixture using ultrasonicator. Carbon black acetylene was added as an additive in the catalyst to separate the catalyst layer with the ratio of catalyst to carbon black was 8 : 2. The loading of the catalyst on the GC surface was about 0.05 mg, which corresponds to normalized loading of 0.714 mg cm^{-2}. Similarly, IrO_2 obtained from Sigma Aldrich was also used to modify the electrode using same procedure for comparative study. All the potentials were converted to reversible hydrogen electrode (RHE) by using the following eqn (1):[41]

$$E_{RHE} = E_{Hg/HgO/OH^-} + 0.118 + (0.059\ pH) \quad (1)$$

The electrochemical measurements were performed on MetrohmAutolab 204 N instrument. The current density was normalized according to geometric surface area and electrochemical surface area. The overpotential (η) was calculated by subtracting the standard potential of water oxidation (1.23 V *vs.* RHE) from the experimental potential observed. The Double layer capacitance (C_{dl}) calculated from the slope of the plot of current density *vs.* scan rate, was used to estimate the electrochemically active surface area (ECSA) using eqn (2).

$$ECSA = C_{dl}/C_s \quad (2)$$

where C_s, is the specific capacitance of the atomically smooth surface of the material and is taken to be 0.04 mF cm^{-2}.[42]

The Tafel slope was calculated by using the eqn (3):

$$\eta = b \log j + a \quad (3)$$

where η is the overpotential and b is the Tafel slope. The electrochemical impedance spectroscopy (EIS) measurements were performed by applying AC potential with 10 mV amplitude in a frequency range of 0.1 Hz to 10^5 Hz at 1.62 V *vs.* RHE. The catalyst stability was analyzed by drop casting the catalyst material on glassy carbon plate electrode in 1 M KOH for 14 hours at a constant potential 1.65 V *vs.* RHE. After the electrochemical stability, the spent catalyst material was used for post catalytic characterizations *i.e.*, FTIR, Raman, SEM, EDX and XPS.

Calculation of TOF

The values of TOF were evaluated by assuming all the metal ions as the active catalytic centers as per the following eqn (4).[43]

$$TOF = J \times A/4 \times F \times n \quad (4)$$

Here, J, A, F and n are the observed current density, geometrical surface area of the working electrode (0.07 cm^2), Faraday

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

constant (96 485 C mol^{-1}) and number of moles of active sites present in the catalyst. Since four electrons are needed to evolve one mole of O_2, so the term $\frac{1}{4}$ is used. Further, n was determined by dividing the mass of material loaded (0.05 mg) onto the electrode surface to the molecular mass of the sample (eqn (5)).

$$n = m/M \quad (5)$$

$$n = \frac{0.05 \times 10^{-3}\ \text{g}}{413\ \text{g mol}^{-1}} = 1.21 \times 10^{-7}\ \text{moles}$$

Results and discussions

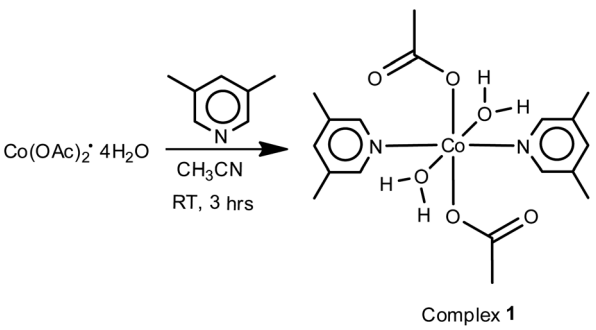
Synthesis

The schematic diagram for synthetic details of complex **1** is given in Scheme 1. The reactions of $Co(OAc)_2 \cdot 4H_2O$ with 3,5-lutidine in 1 : 2 ratio in acetonitrile at room temperature afforded complex **1** in quantitative yields. IR spectra of complex **1** (Fig. SI-3a in ESI†), shows bands at 1572 (s) and 1419 (s) for ν_{asym}(OCO) and ν_{sym}(OCO) characteristic for monodentate acetate moieties. The broad band at 3425 cm^{-1} confirms the presence of coordinated water molecules. Further, the bands at $\rho r(H_2O)$ (710 cm^{-1}) and $\rho w(H_2O)$ (649 cm^{-1}) also support the presence of coordinated water molecules.[44]

Crystal structure

The molecular structure of complex **1** was determined by single-crystal X-ray diffraction data. The molecular structure of complex **1** with the atom labeling scheme is shown in Fig. 1. Complex **1** consists of a Co(II) atom coordinated by two 3,5-lutidine, two water molecules and two acetate moieties in monodentate mode to afford an octahedral geometry. The $Co1–O1_{water}$, $Co1–O2_{acetate}$ and $Co1–N1_{lutidine}$ distances in complex **1** are 2.131(1), 2.068(1) and 2.153(2) Å, respectively.

Significant hydrogen-bond parameters observed in complex **1** are listed in Table 1. The oxygen atom (O1) of water molecules in **1** act as hydrogen-bond donor, one to the carbonyl oxygen (O3) of the monodentate acetate within the molecule (intramolecular hydrogen bonding) and another with the carbonyl oxygen (O3) of the monodentate acetate of adjacent molecule



Scheme 1 Synthesis scheme of complex **1**.

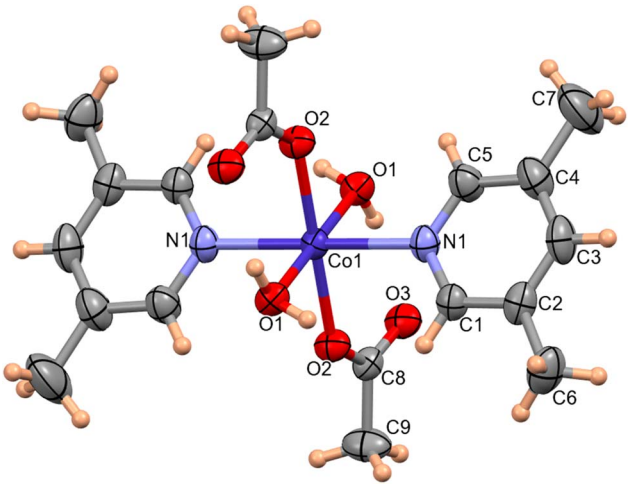


Fig. 1 ORTEP of complex **1** with 50% probability ellipsoid. Selected bond lengths (Å) and angles (°): Co1–O1 2.133(1), Co1–O2 2.068(1), Co1–N1 2.153(2); O1–Co1–O1 180.00(0), O2–Co1–O2 180.00(6), O2–Co1–N1–90.58(5), N1–Co1–N1 180.00(8), O2–Co1–O1 88.55(4), N1–Co1–O1 87.42(5).

(intermolecular hydrogen bonding) as depicted in Fig. 2a. The aforementioned hydrogen-bonding network grows to afford a three-dimensional supramolecular packing along the *c*-axis, as depicted in Fig. 2b. From packing diagram of complex **1**, it was also observed that the crystal structure has various hexagonal vacant pores formed by the supramolecular assembly which could be useful for enhancing their catalytic, separation and gas storage properties. The void volume has also been calculated to be 114 $Å^3$ per unit cell. Although the structural analysis shows a solvent accessible void volume of around 194 $Å^3$ per unit cell, it is not apparent that water molecules can enter this void and the residual density value (0.7 Å) also do not imply the presence of water molecules inside this channel. The porous nature of complex **1** is also clearly visualized in FESEM image as depicted in Fig. SI-1 in ESI.†

Computational studies

Frontier molecular orbitals. The optimized geometry of complex **1** is depicted in Fig. SI-2.† The HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) along with HOMO−1 and LUMO+1 have been computed through DFT using GaussView 5.0 software. Generally, the electron donating orbital is regarded as HOMO and electron accepting is LUMO. These two orbitals are most interacting in the complex, and thus called frontier orbitals.[45–48] The energy gaps were computed at LANL2DZ basis set. In

Table 1 Significant hydrogen bond distances (Å) and bond angles (°) in the crystal structure of complex **1**

D–H···A	D–H	H···A	D···A	∠D–H···A
O1–H1A···O3	0.860	1.946	2.708(2)	148.70
O1–H1B···O3	0.861	2.080	2.823(2)	148.84

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

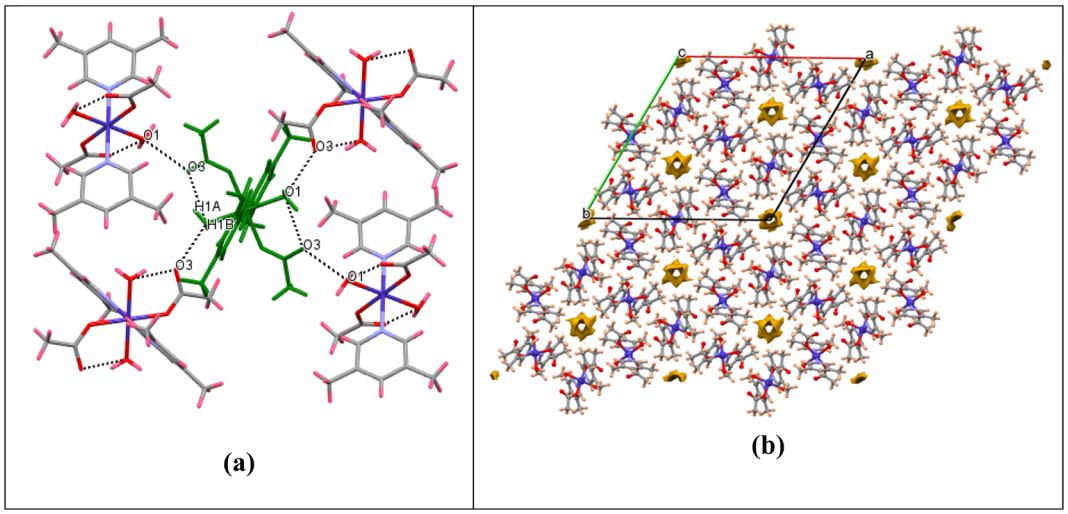


Fig. 2 (a) Packing diagram showing O–H···O hydrogen bonding for complex **1**. (b) Packing diagram showing the void surfaces in the crystal structure of complex **1**, viewed down crystallographic *c*-axis.

complex **1**, the HOMO and LUMO energies are −5.19 and −1.27 eV, respectively, leading to an energy gap of 3.92 eV (Fig. 3). Determining the energy gap is important because it displays the charge transfer behavior in a system,[49] and the chemical reactivity and stability of the molecule as a whole depends largely on the energy gap. Larger the HOMO–LUMO energy gap, greater the stability and lower is the chemical reactivity of the molecule and *vice versa*.[50,51] From the frontier molecular orbital (FMO) diagram, it has been observed that the electron density is densely accumulated around the metal ion and oxygen atoms in HOMO, while it is dispersed on and around the pyridine rings in LUMO. A slightly more gathered electronic density at the metal center is observed in HOMO−1. Thus, it is believed that the metal center is behaving as an active site for the incoming substrate to get adsorbed/attached. The subsequent steps for OER mechanism are expected to proceed from here.

Molecular electrostatic potential surfaces. The molecular electrostatic potential (MEP) maps provide information about the relative polarity of a particular system. The MEP of complex **1** (Fig. 4) has been computed through DFT mode, using B3LYP level of theory and LANL2DZ as basis set. The scale having different colors here indicates varied electrostatic potentials within a molecular system.[52] The order of potentials from low to high goes as red < orange < yellow < green < blue. On one end of the scale, the blue color suggests the presence of electropositive potential, while on the other end, the red color indicates the electronegative potentials, and the green color specifies zero potential. In this way, we can understand the relative polarity of a particular system.[53,54] In the present case, we observe a deep red color at the oxygen centers of acetate ligand while a positive potential around the substituted pyridine ring. Lone pair of electrons is generally associated with a negative potential; thus, a reddish coloration is observed around the oxygen centers. In addition to this, the yellowish coloration around the metal

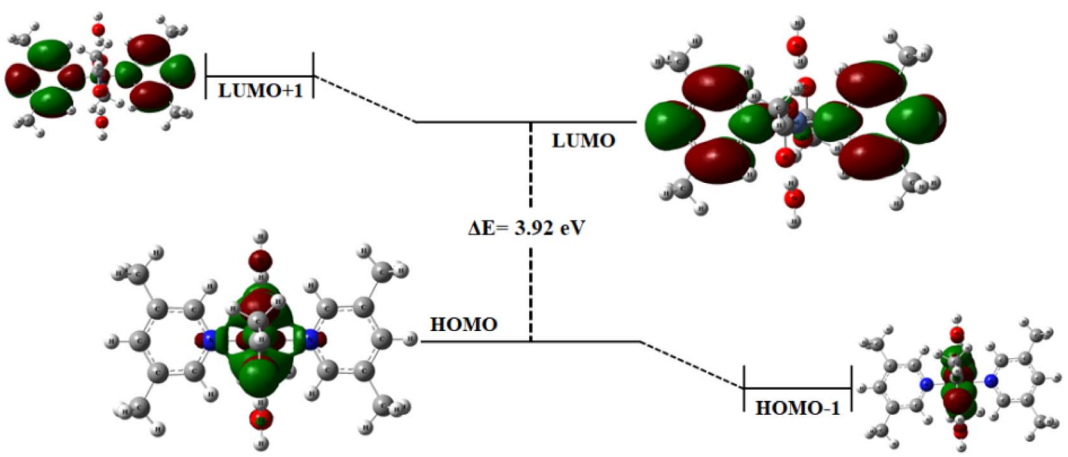


Fig. 3 Frontier molecular orbital diagram for complex **1** at iso-surface value = 0.02.

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

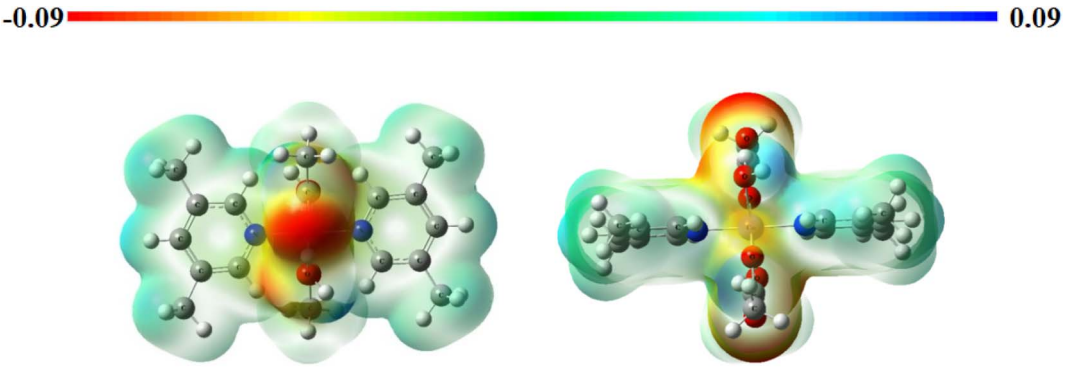


Fig. 4 Molecular electrostatic potential maps for complex **1** (iso-surface value = 0.002).

center depicts the presence of electron density in this region. These observations can be well correlated with that of FMOs, as discussed above.

Non-covalent interactions. The reduced gradient of density (RDG) plots reveals information about weak, intramolecular, and intermolecular interactions.[55,56] E. R. Johnson first introduced the RDG function, which is density and its first derivative together, representing the divergence from homogeneous electronic distribution.[57] It is a dimensionless quantity. Based on the reduced density gradient, s and electron density, ρ, the NCI index plots are drawn as per eqn (6)

$$s = 1/(2(3\pi^2)^{1/3}) \times |\nabla\rho|/\rho^{4/3} \quad (6)$$

The kind of interactions existing within a molecule is further understood by the Laplacian of density, $\nabla^2\rho$, which is the sum of three eigenvectors λ_1, λ_2 and λ_3. In a certain system, λ_3 differs along the internuclear direction, whereas the other two components (*viz.* λ_1 and λ_2) describe the variation of density in the plane normal to that of λ_3 eigenvector. The value of eigenvector λ_2 mainly suggests the type of interaction present in a system. For bonding interactions, such as H-bond, the value of $\lambda_2 < 0$, while for non-bonding interactions λ_2 is greater than zero. The value of λ_2 is nearly zero for weak van der Waals interactions.[57,58] The colors in RDG plot also suggest these interactions visually. For example, the blue color stands for H-bonds, green color suggests the presence of weak van der Waals interactions and the red color depicts the steric effect.[54] Fig. 5 represents the non-covalent interactions in complex **1** and their two-dimensional RDG plots. In complex **1**, the RDG plot displays the presence of H-bonds and steric repulsions co-existing within the complex.

Oxygen evolution activity in alkaline electrolyte

The electrocatalytic activity of the as prepared complex **1** was evaluated using Linear Sweep voltammetry (LSV) at scan rate 10 mV s^{-1} in 1 M KOH solution. The complex **1** showed onset potential of 1.50 V *vs.* RHE and high current density of 100 mA cm^{-2} at 1.87 V *vs.* RHE. The observed results towards OER electrocatalysis are even better than the state-of-the-art IrO_2 catalyst which showed OER onset potential of 1.52 V *i.e.*, 20 mV higher than complex **1**. The comparative LSV is shown in Fig. 6a. The kinetics of the catalysis towards OER of the complex **1** was also compared to that of IrO_2 catalyst using corresponding Tafel slope values. The commercial IrO_2 exhibited a slope value of 64.6 mV dec^{-1}, whereas the complex **1** showed 47.7 mV dec^{-1}. Tafel slope indicating that our catalyst is showing facile kinetics towards OER, higher than commercial IrO_2 catalyst (Fig. 6b). The complex **1** showed an overpotential of 530 mV to reach 40 mA cm^{-2} which is 70 mV lower than that of IrO_2 (Fig. 6c). The comparison with similar materials in literature

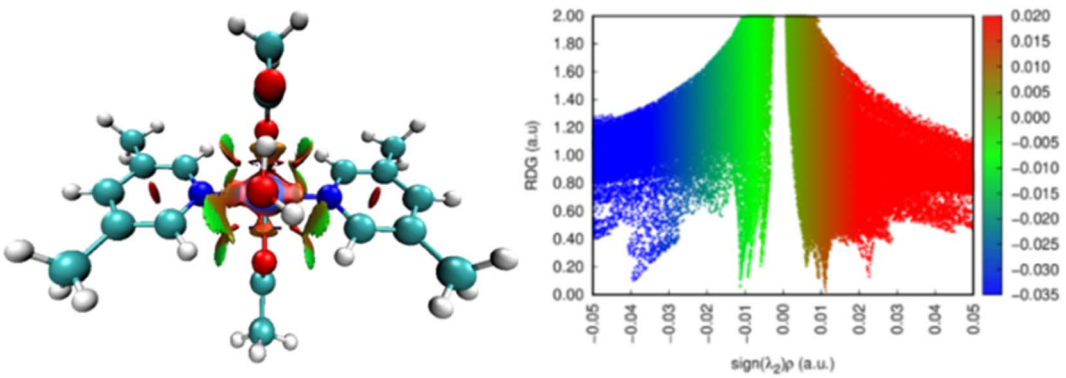


Fig. 5 Non-covalent interactions and RDG plots for complex **1** (iso-surface value = 0.5).

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

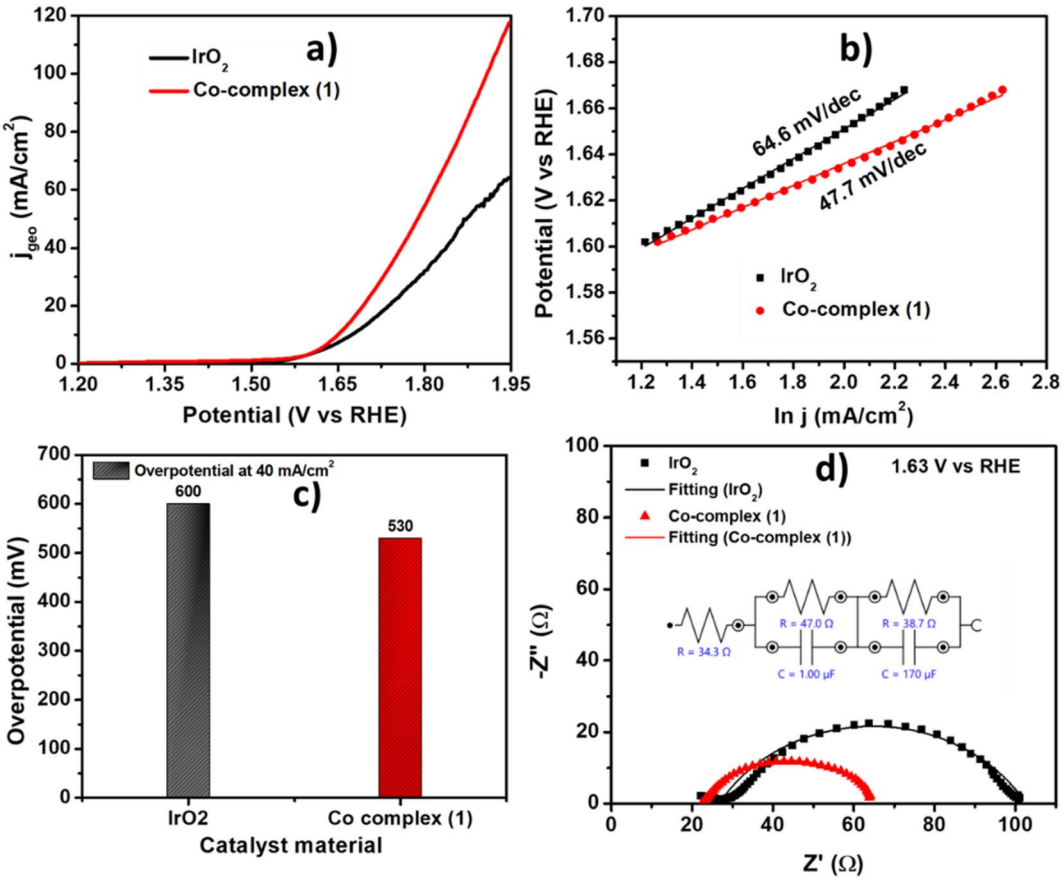


Fig. 6 (a) Comparison of LSV polarization curve for complex **1** and IrO_2 at 10 mV s^{-1} in 1 M KOH, (b) the corresponding Tafel plots, (c) the overpotential comparison at 40 mA cm^{-2} and (d) the Nyquist plot comparison for complex **1** and IrO_2.

was also shown in Table 2 indicating the appreciable activity of our catalyst towards OER.[34,36–38,59–61] The electrochemical impedance spectroscopy (EIS) was performed at a potential of 1.63 V *vs.* RHE to determine the charge transfer parameter for complex **1**. The diameter of the Nyquist plot is related to the charge transfer resistance at the electrode–electrolyte interface. Nyquist plots shown in Fig. 6d, showed that the complex **1** has low charge transfer resistance of 43.5 Ω, also lower than IrO_2 showing R_{ct} of 63.8 Ω which is also consistent with the lower overpotentials required to reach higher current density from the LSV and Tafel analysis.

The C_{dl} was calculated by performing CV at different scan rates (Fig. 7a and b) and plotting the current density *vs.* scan rate (Fig. 7c) which was found to be 81.3 μF and 75.8 μF for complex **1** and IrO_2 complex respectively. The current density normalized with ECSA was also shown in Fig. 7d, indicating the good electrocatalytic activity of Co-complex **1** in terms of high

Table 2 Comparison of the efficacy molecular cobalt electrocatalyst for OER

Cobalt complexes	Electrolyte	Overpotential (mV)/current density (mA cm^{-2})	Tafel slope (mV dec^{-1})	TOF	Ref.
$[(Lut)_2Co(OAc)_2(H_2O)_2]$	1 M KOH	410 ± 20/10.0	45.7	0.05 s^{-1}	This work
$[Co(LN_2O_3)H_2O]$	Sodium borate (pH = 11)	500	—	—	34
$[(L_{monoanionic})_2Co(CH_3OH)_4]^a$	Phosphate buffer (pH = 9)	520/0.5	—	5 s^{-1}	36
$[Co\{(SeP^iPr_2)_2N\}_2$	1 M KOH	320/10.0	61.6	0.032 s^{-1}	37
$[Co(bipyalk)(OAc)_2]^a$	Phosphate buffer (pH = 6)	360/1.0	—	1.5 s^{-1}	38
$[Co@CB[5]]^a$/ITO	Borate buffer (pH = 9.2)	485/1.0	59.5	0.3 s^{-1}	59
$[Co^{II}Co^{III}(\mu\text{-}OAc)(\mu_3\text{-}OH)(\mu\text{-}L_{trianionic})]_2^a$	Sodium phosphate buffer of pH 7	768/1.0	320	1.1 × 10^{-3} h^{-1}	60
$[CoL_{dianionic}]^a$	Buffer pH-11	360/10.0	135.9	—	61

[a] $L_{monoanionic}$ = (*E*)-4-((2-hydroxynaphthalen-1-yl)diazenyl)benzenesulfonate), bipyalk = 2,2′-([2,2′-bipyridine]-6,6′-diyl)bis(propan-2-ol), CB[5] = cucurbit[5]uril, $L_{trianionic}$ = *N*′,*N*″-(5-methyl-2-oxido-1,3-phenylene)bis(methan-1-yl-1-ylidene)bis(4-methoxybenzoylhydrazonate), $L_{dianionic}$ = 4-chloro-1,2-bis[2-hydroxy-5-(phenylazo)benzylideneamino]benzene).

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

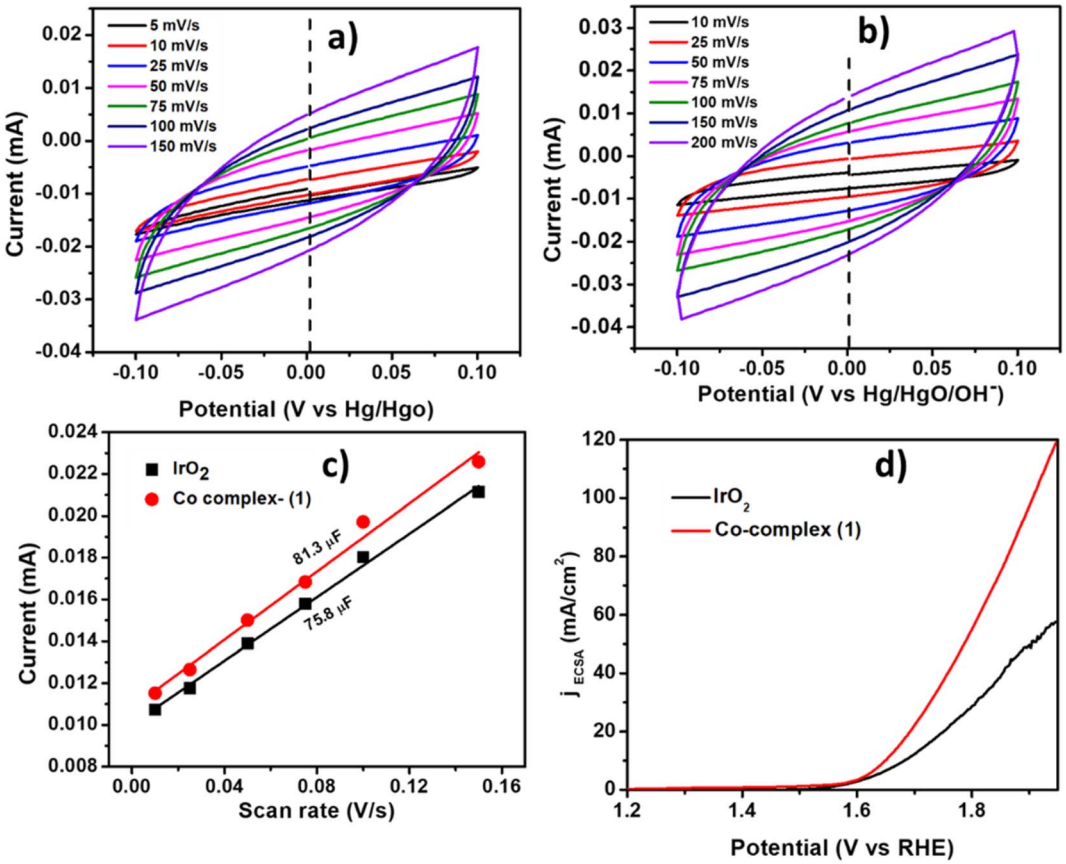


Fig. 7 The CVs (cyclic voltammograms) at different scan rates in the capacitive region for (a) IrO_2 and (b) complex **1**, (c) the plot of current *vs.* scan rates for C_{dl} calculation, (d) the comparison of LSV curves normalized by ECSA for complex **1** and IrO_2.

current density towards electrocatalytic OER. The summary of the OER electrochemical parameters is shown in Table 3.

The turn over frequency, which indicates the intrinsic activity of the catalyst was calculated by using the equation given in the experimental section. The complex **1** showed the two times higher TOF ($0.05\ s^{-1}$) at 1.73 V as compared to IrO_2 ($0.02\ s^{-1}$) (Fig. 8a). The long term stability of complex **1** was also checked using chronoamperometry method, by applying a constant potential of 1.65 V for 14 hours. It was observed (Fig. 8b) that the catalyst showed high stability with retention of 95% current. The CV curves before and after the OER stability are given in the inset of Fig. 8b, showing a slight decrement in

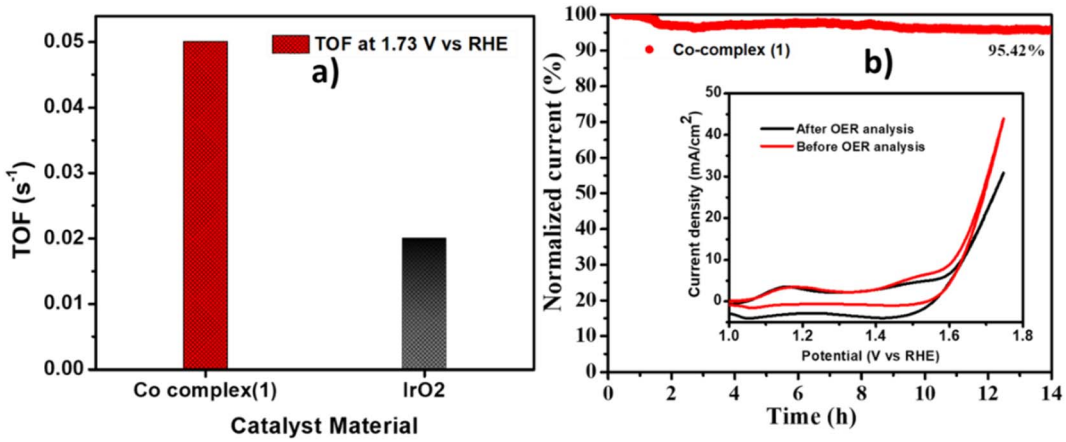


Fig. 8 (a) The comparison of TOF at 1.73 V *vs.* RHE for Co-complex (**1**) and IrO_2, (b) chronoamperometry curve at 1.65 V *vs.* RHE for 14 hours for Co-complex (**1**).

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

Table 3 The comparison of activity parameters for OER for Co-complex (**1**) and IrO_2

Parameters	Complex **1**	IrO_2
E_{onset} (V *vs.* RHE)	1.51 ± 0.01	1.52
η_{10mA/cm^2} (mV *vs.* RHE)	410 ± 20	440
Tafel slope (mV dec^{-1})	45.7	64.6
Mass activity (mA mg^{-1}) @ 1.64	14	8.5
R_{ct} (Ω)	43.5	63.8
C_{dl} (μF)	81.3	78.5
TOF (1.73 V *vs.* RHE)	0.05	0.02
α	0.59	0.56

the current density after 14 hours of stability, indicating robustness of the catalyst.

The post catalytic characterisation of the spent electrodes (after OER analysis) were performed after the durability test on the Co-complex (**1**). The post catalytic FTIR (Fig. SI-3a in ESI†) shows the peak at 3431 cm^{-1} corresponding to the –OH vibration, 2921 cm^{-1} for Co–OH stretching. The peak present at 1633 cm^{-1} corresponds to Co–O.[62,63] These peaks in the post OER FTIR suggests that the formation of $Co(OH)_2$/CoOOH could act as the main active sites for the OER. However, along with these, some peaks matching with the initial spectra at 1038 and 1153 cm^{-1} indicate that the complex **1** is not totally degraded/decomposed after catalysis. The spectra suggests that the oxidation of Co(II) to Co(III) occurs at high oxidation potential and Co act as the catalytic centre for OER. The FESEM image after catalysis shows agglomeration of the particles (Fig. SI-3b in ESI†). The EDX spectra as shown in Fig. SI-4a in ESI†) indicates the presence of C, N, Co and O in the Co-complex (**1**). After catalysis (Fig. SI-4b in ESI†), the EDX of the sample displays an increase in the O content as well as presence of Co, C, N and K (due to KOH as electrolyte) indicating the presence of $Co(OH)_2$ at the surface of the catalyst. The elemental mapping (Fig. SI-5 in ESI†) shows the uniform distribution of the elements on the surface before and after catalysis which indicates that the complex is not totally changed to hydroxide species. This suggests that formation of thin layer of $Co(OH)_2$/CoOOH can be the reason for high activity of Co-complex (**1**) towards OER.

The X-ray photoelectron spectroscopy was performed to further understand the chemical composition and oxidation state in the pre and post catalytic OER samples as shown in Fig. 9. The Co 2p core level spectra of Co-complex (**1**) (Fig. 9a) shows two peak at 780.85 eV (Co^{2+} $2p_{3/2}$) and 796.49 eV (Co^{2+} $2p_{1/2}$) and two strong satellite peaks confirming the presence of Co^{2+} in the sample. After OER analysis, the slight shift of the Co^{2+} peak to 780.19 eV and the appearance of Co^{3+} peak at 782.18 eV is in accordance with the FTIR analysis, confirming that the formation of $Co(OH)_2$ and CoOOH at the surface.[64] The M–O and M–OH peak intensity increases in the post catalytic O 1s spectra (Fig. 9b) also showing the formation of $Co(OH)_2$ at the surface. The C 1s XPS spectra shows the presence of C–C

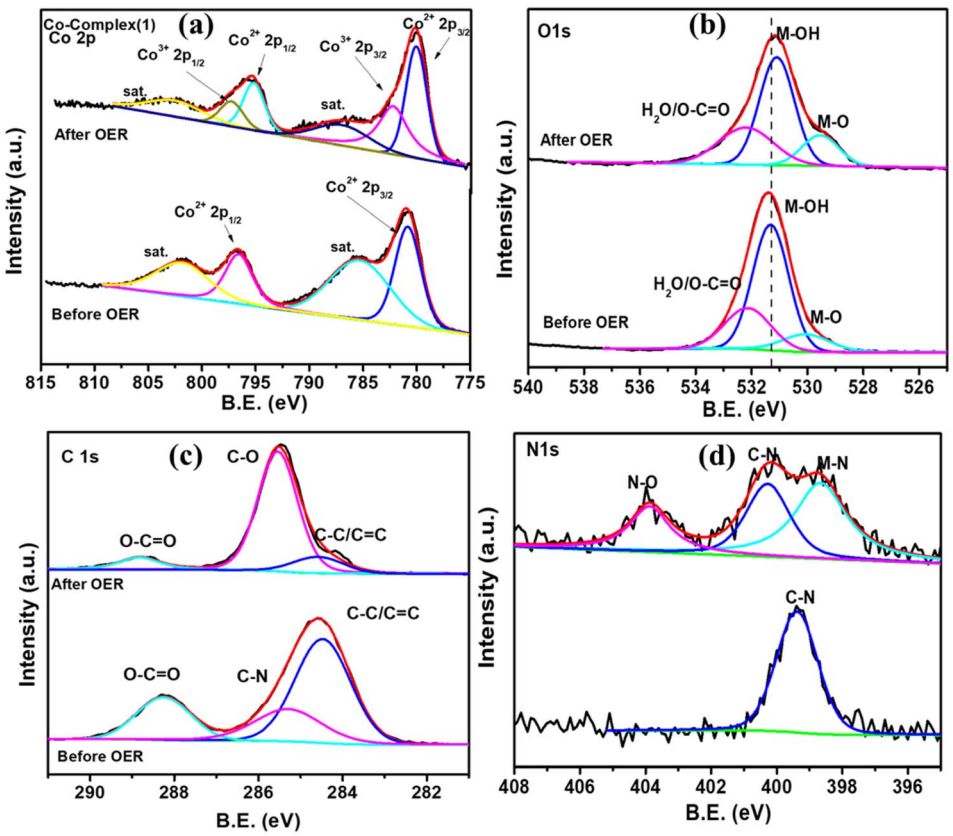


Fig. 9 XPS core level spectra comparison of (a) Co 2p, (b) O 1s, (c) C 1s and (d) N 1s before and after catalysis of Co-complex (**1**).

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

(284.6), C–N (285.4) and O–C=O (288.2) in the pristine sample. After catalysis, the relative composition of C–C decreases and C–O increases. This indicates the oxidation of C due to the oxidative environment provided by OER. Similarly the occurance of N–O peak in N 1s spectra is also due to oxidation. Thus, the post catalytic XPS, FTIR and EDX analysis confirmed the retention of the electrocatalyst in terms of bulk composition, and formation of Co^{2+} ($Co(OH)_2$) and Co^{3+} (CoOOH) plays a crucial role during OER.

The high OER performance of the Co-complex (**1**) may have been resulted due to its highly porous connected network, high accessible Co active centres and high surface area. The interconnected network arising from the OH···O hydrogen bonding leads to a 3-D structure with hexagonal vacant pores which helped inenhancing mass transfer and ion transport at the electrode–electrolyte interface leading to high electrocatalytic performance of the Co-complex **1**.

Conclusion

We have the novel lutidine coordinated cobalt(II) acetate complex $[(3,5\text{-lutidine})_2Co(OAc)_2(H_2O)_2]$ (**1**) by simplest condensation reaction. The structure of complex **1** was determined by single crystal X-ray diffraction data and found distorted octahedral geometry around the metal center. The noncovalent interaction found in complex **1** resulted in O–H···O hydrogen-bonding network that grown in a three-dimensional structure with various hexagonal vacant pore which could be useful for enhancing their catalytic activity. The comparative study of geometrical parameters obtained from the XRD data and DFT calculations was found in good agreement. The MEP maps analysis depicted the electron rich centers in the complex **1** are also found well in agreement with the solid-state structure. It was thought that the coordination flexibility and electron rich centers around metal center will also support in the enhancement of their catalytic activity. The Co-complex **1** showed excellent electrocatalytic OER activity better than state-of-the-art IrO_2, which may be due to the 3D porous structure, rapid transport of ions and high accessible Co active sites, higher electron density, and coordination flexibility around metal center. Further, the post catalytic XPS, FTIR and EDX analysis also confirmed the retention of the electrocatalyst in terms of bulk composition, and formation of Co^{2+} ($Co(OH)_2$) and Co^{3+} (CoOOH) OER.

Data availability

All data generated or analyzed during this study are included in this published article and its ESI.†.

Author contributions

DT: investigation, writing, PJ: electrochemical study and writing, DA: computational studies, Athul Sudheendranath and Sajesh P. Thomas: SCXRD, UK: supervision, writing, review & editing, PPI: supervision, writing, review & editing.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

The authors UK and PPI acknowledge the Science and Engineering Board (SERB) New Delhi, India for funding [TAR/2020/000232]. DT and PJ acknowledge University Grant Commission and IIT Delhi for research fellowship. Authors also acknowledge the Department of Chemistry and University Science Instrumentation Center, University of Delhi, and Department of Chemistry, IIT Delhi for instrumental facilities.

References

1 T. Yao, X. An, H. Han, J. Q. Chen and C. Li, *Adv. Energy Mater.*, 2018, **8**, 1800210.
2 D. Fa, J. Yuan, G. Feng, S. Lei and W. Hu, *Angew. Chem., Int. Ed.*, 2023, **62**, e202300532.
3 Y. Qiu, Z. Liu, X. Zhang, A. Sun and J. Liu, *J. Colloid Interface Sci.*, 2022, **625**, 50–58.
4 Y. Chen, M. Wang, S. Xiang, J. Liu, S. Feng, C. Wang, N. Zhang, T. Feng, M. Yang, K. Zhang and B. Yang, *ACS Sustainable Chem. Eng.*, 2019, **7**, 10912–10919.
5 T. Ghosh and G. Maayan, *Angew. Chem., Int. Ed.*, 2019, **58**, 2785–2790.
6 S. W. Gersten, G. J. Samuels and T. J. Meyer, *J. Am. Chem. Soc.*, 1982, **104**, 4029–4030.
7 T. Reier, M. Oezaslan and P. Strasser, *ACS Catal.*, 2012, **2**, 1765–1772.
8 L. C. Seitz, C. F. Dickens, K. Nishio, Y. Hikita, J. Montoya, A. Doyle, C. Kirk, A. Vojvodic, H. Y. Hwang, J. K. Norskov and T. F. Jaramillo, *Science*, 2016, **353**, 1011–1014.
9 Q. Shi, C. Zhu, D. Duand and Y. Lin, *Chem. Soc. Rev.*, 2019, **48**, 3181–3192.
10 B. Deng, Y. Long, C. Yang, P. Du, R. Wang, K. Huang and H. Wu, *Chem. Commun.*, 2021, **57**, 7830–7833.
11 S. Anantharaj, S. R. Ede, K. Sakthikumar, K. Karthick, S. Mishra and S. Kundu, *ACS Catal.*, 2016, **6**, 8069–8097.
12 Q. Liang, J. Chen, F. Wang and Y. Li, *Coord. Chem. Rev.*, 2020, **424**, 213488.
13 J. Yu, F. A. Garcés-Pineda, J. González-Cobos, M. Peña-Díaz, C. Rogero, S. Giménez, M. C. Spadaro, J. Arbiol, S. Barja and J. R. Galán-Mascarós, *Nat. Commun.*, 2022, **13**, 4341.
14 Z. P. Wu, X. F. Lu, S. Q. Zang and X. W. Lou, *Adv. Funct. Mater.*, 2020, **30**, 1910274.
15 W. Zhang, L. Cui and J. Liu, *J. Alloys Compd.*, 2020, **821**, 153542.
16 A. Badruzzaman, A. Yuda, A. Ashok and A. Kumar, *Inorg. Chim. Acta*, 2020, **511**, 119854.
17 L. Reith, C. A. Triana, F. Pazoki, M. Amiri, M. Nyman and G. R. Patzke, *J. Am. Chem. Soc.*, 2021, **143**, 15022–15038.
18 B. Paul, P. Bhanja, S. Sharma, Y. Yamauchi, Z. A. Alothman, Z. L. Wang and R. B. A. Bhaumik, *J. Colloid Interface Sci.*, 2021, **582**, 322–332.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Open Access Article. Published on 15 August 2023. Downloaded on 9/5/2023 10:24:55 AM. This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

19 F. T. Haase, A. Bergmann, T. E. Jones, J. Timoshenko, A. Herzog, H. S. Jeon, C. Rettenmaier and B. R. Cuenya, *Nat. Energy*, 2022, **7**, 765–773.
20 Y. Zhang, Q. Fu, B. Song and P. Xu, *Acc. Mater. Res.*, 2022, **3**, 1088–1100.
21 F. Lyu, Y. Bai, Q. Wang, L. Wang, X. Zhanga and Y. Yin, *Dalton Trans.*, 2017, **46**, 10545–10548.
22 B. Zhang, J. Zhang, X. Tan, D. Tan, J. Shi, F. Zhang, L. Liu, Z. Su, B. Han, L. Zheng and J. Zhang, *Chem. Commun.*, 2018, **54**, 4045.
23 S. Y. Chae, Y. J. Hwang, J. H. Choi and O. S. Joo, *Electrochim. Acta*, 2013, **114**, 745–749.
24 S. B. Kale, A. C. Lokhande, R. B. Pujari and C. D. Lokhande, *J. Colloid Interface Sci.*, 2018, **532**, 491–499.
25 Y. Kang, Y. He, D. Pohl, B. Rellinghaus, D. Chen, M. Schmidt, V. Süß, Q. Mu, F. Li, Q. Yang, H. Chen, Y. Ma, G. Auffermann and G. L. C. Felser, *ACS Appl. Mater. Interfaces*, 2022, **14**, 19324–19331.
26 M. Liaoa, G. Zenga, T. Luoa, Z. Jina, Y. Wanga, X. Koua and D. Xiao, *Electrochim. Acta*, 2016, **194**, 59–66.
27 Q. S. Zhong, W. Y. Xia, B. C. Liu, C. W. Xu and N. Li, *Int. J. Hydrogen Energy*, 2019, **44**, 10182–10189.
28 J. Chang, Y. Xiao, M. Xiao, J. Ge, C. Liu and W. Xing, *ACS Catal.*, 2015, **5**, 6874–6878.
29 M. Zhu, Y. Zhou, Y. Sun, C. Zhu, L. Hu, J. Gao, H. Huang, Y. Liu and Z. Kang, *Dalton Trans.*, 2018, **47**, 5459–5464.
30 P. Chen, K. Xu, Y. Tong, X. Li, S. Tao, Z. Fang, W. Chu, X. Wua and C. Wu, *Inorg. Chem. Front.*, 2016, **3**, 236–242.
31 X. Peng, C. Pi, X. Zhang, S. Li, K. Huo and P. K. Chu, *Sustainable Energy Fuels*, 2019, **3**, 366.
32 D. K. Dogutan, R. McGuire Jr and D. G. Nocera, *J. Am. Chem. Soc.*, 2011, **133**, 9178–9180.
33 D. Wang and J. T. Groves, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 15579.
34 S. Gonawala, H. Baydoun, L. Wickramasinghe and C. N. Verani, *Chem. Commun.*, 2016, **52**, 8440–8443.
35 Q. Daniel, R. B. Ambre, B. Zhang, B. Philippe, H. Chen, F. Li, K. Fan, S. Ahmadi, H. Rensmo and L. Sun, *ACS Catal.*, 2017, **7**, 1143–1149.
36 H. T. Shi, X. X. Li, F. H. Wu and W. B. Yu, *Dalton Trans.*, 2017, **46**, 16321–16326.
37 I. M. Abdullahi, J. Masud, P. C. Ioannou, E. Ferentinos, P. Kyritsis and M. Nath, *Molecules*, 2021, **26**, 945.
38 Y. F. Su, W. Z. Luo, W. Q. Lin, Y. B. Su, Z. J. Li, Y. J. Yuan, J. F. Li, G. H. Chen, Z. Li, Z. T. Yu and Z. Zou, *Angew. Chem., Int. Ed.*, 2022, **61**, e202201430.
39 A. Valizadeh, R. Bikas, S. Nandy, T. Lis, K. H. Chae and M. M. Najafpour, *Dalton Trans.*, 2022, **51**, 220–230.
40 U. Kumar, J. Thomas and N. Thirupathi, *Inorg. Chem.*, 2010, **49**, 62–72.
41 P. Jain, S. Jha and P. P. Ingole, *Sustainable Energy Fuels*, 2022, **6**, 1094–1107.
42 Y. Yang, H. Fei, G. Ruan and J. M. Tour, *Adv. Mater.*, 2015, **27**, 3175–3180.
43 R. K. Tripathy, A. K. Samantaraand and J. N. Behera, *Dalton Trans.*, 2019, **48**, 10557–10564.
44 S. C. Manna, S. Mistria and A. D. Jana, *CrystEngComm*, 2012, **14**, 7415–7422.
45 C. Quintana, G. Silva, A. H. Klahn, V. Artigas, M. Fuentealba, C. Biot, I. Halloum, L. Kremer, N. Novoa and R. Arancibia, *Polyhedron*, 2017, **134**, 166–172.
46 C. Soriano-Correa, C. Barrientos-Salcedo, M. Francisco-Marquez and C. I. Sainz-Díaz, *J. Mol. Graphics Modell.*, 2018, **81**, 116–124.
47 N. Elangovan, R. Thomas, S. Sowrirajan, K. P. Manoj and A. Irfan, *Polycyclic Aromat. Compd.*, 2021, **42**, 6818–6835.
48 S. M. Hiremath, A. Suvitha, N. R. Patil, C. S. Hiremath, S. S. Khemalapure, S. K. Pattanayak, V. S. Negalurmath and K. Obelannavar, *J. Mol. Struct.*, 2018, **1171**, 362–374.
49 K. Karrouchi, S. A. Brandán, Y. Sert, H. El-marzouqi, S. Radi, M. Ferbinteanu, M. E. A. Faouzi, Y. Garcia and M. Ansa, *J. Mol. Struct.*, 2020, **1219**, 128541.
50 M. Alizadeh, Z. Mirjafary and H. Saeidian, *J. Mol. Struct.*, 2020, **1203**, 127405.
51 N. Elangovan, R. Thomas and S. Sowrirajan, *J. Mol. Struct.*, 2022, **1250**, 131762.
52 M. Shahid, M. Salim, M. Khalid, M. N. Tahir, M. U. Khan and A. A. C. Braga, *J. Mol. Struct.*, 2018, **1161**, 66–75.
53 M. Thirunavukkarasu, G. Balaji, S. Muthu, B. R. Raajaraman and P. Ramesh, *Chem. Data Collect.*, 2021, **31**, 100622.
54 K. P. Manoj, N. Elangovan and S. Chandrasekar, *Inorg. Chem. Commun.*, 2022, **139**, 109324.
55 A. Prabakaran, V. Vijayakumar, N. Radhakrishnan, R. Chidambaram and S. Muthu, *Polycycl. Aromat. Compd.*, 2020, **42**, 925–941.
56 J. M. del Campo, J. L. Gázquez, R. J. Alvarez-Mendez and A. Vela, *Int. J. Quantum Chem.*, 2012, **112**, 3594–3598.
57 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498–6506.
58 C. Lefebvre, G. Rubez, H. Khartabil, J. C. Boisson, J. C. García and E. Hénon, *Phys. Chem. Chem. Phys.*, 2017, **19**, 17928–17936.
59 F. Li, H. Yang, Q. Zhuo, D. Zhou, X. Wu, P. Zhang, Z. Yao and L. Sun, *Angew. Chem.*, 2021, **133**, 2004–2013.
60 A. K. Srivastava, A. Mondal, S. Konar and S. Pal, *Dalton Trans.*, 2022, **51**, 4510–4521.
61 Z. Shaghaghi, P. S. Kouhsangini and R. Mohammad-Rezaei, *Appl. Organomet. Chem.*, 2021, **35**, e6103.
62 A. D. Jagadale, D. P. Dubal and C. D. Lokhande, *Mater. Res. Bull.*, 2012, **47**, 672–676.
63 M. Adeel, S. Parisi, M. Mauceri, K. Asif, M. Bartoletti, F. Puglisi, I. Caligiuri, M. M. Rahman, V. Canzonieri and F. Rizzolio, *ACS Omega*, 2021, **43**, 28611–28619.
64 P. Jain, S. Jha and P. P. Ingole, *ACS Appl. Energy Mater.*, 2023, **6**, 3278–3290.

© 2023 The Author(s). Published by the Royal Society of Chemistry

Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023)
IEEE Xplore Part Number: CFP23AWO-ART; ISBN: 979-8-3503-9663-8

2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192879

A Review on Community Detection Approach for Recommender Systems

Akansha Mittal
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
akanshamittal2207@gmail.com

Anurag Goel
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
0000-0003-2660-7540

***Abstract*—In the last few years, Recommender systems have become extremely popular for users to extract relevant information. The applicability of recommender systems is in several domains, including e-commerce sites, social media platforms and several others. The Community detection approaches aim to detect the collective users that have similar behavior, interest, and choice. Community detection technique helps in creating a new dimension for recommender systems. This paper presents a thorough review of the Community detection approaches for recommender systems. The challenges and research directions are also discussed in the later part of the paper.**

***Index Terms*—Community Detection, Recommender Systems, Collaborative Filtering, Louvain Algorithm.**

I. Introduction

Recommender systems are algorithms that suggest products, services, or content to customers based on their preferences and past behavior. They are extensively used in many industries, such as entertainment, e-commerce, healthcare, and social media, to give personalized recommendations to users. Recommender systems have become increasingly important due to the upsurge of data and the need to filter and present relevant information to users efficiently. They help users save time and effort by automating the process of finding relevant items from a vast collection of options. Additionally, they help businesses increase user engagement, loyalty, and revenue by improving user experience and providing personalized marketing.

Community detection can be used to recognize clusters of customers with similar preferences and behaviors in a recommender system. By identifying these communities, recommender systems can produce more accurate and diverse recommendations to users. For example, users in the same community may have similar tastes in music, movies, or books and may appreciate the same recommendations. Community detection can also help address the cold start problem, which occurs when a new customer or item has little or no data available for recommendation. By leveraging the community structure, recommender systems can infer the preferences of a new customer based on the preferences of users in the same community. Moreover, community detection can help in enhancing the scalability and interpretability of recommender systems by reducing the dimensionality of the input data and providing meaningful clusters of users and items. Therefore, studying community detection in recommender systems can lead to better recommendation quality, user satisfaction, and business performance.

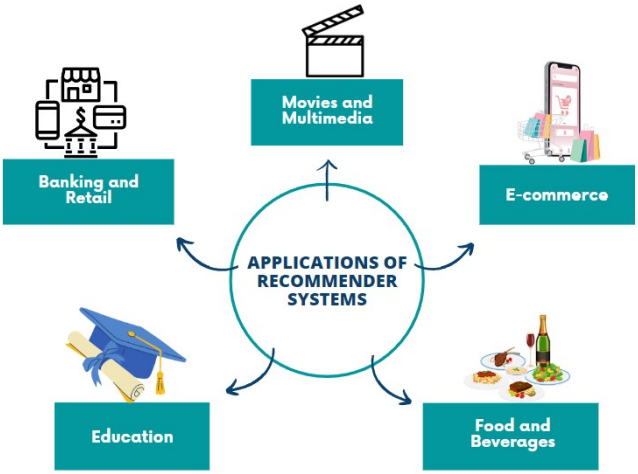


Fig. 1. Applications of Recommender Systems

II. Recommender Systems and Community Detection

A. Types of Recommender Systems

Recommender systems are extensively used in entertainment, e-commerce, social media, and many other fields to suggest items or services of interest to users. These systems can be categorized into several types based on the type of data they use, the algorithms they employ, and the output they generate. Here are some types of recommender systems:

1) Collaborative Filtering (CF): It is a very popular technique in recommender systems that relies on the past behavior of customers and items to make recommendations. This approach can be further divided into user based CF & item based CF [1].
2) Content-Based Recommender System (CBRS): It suggests products based on the similarity between the features of the products and the preferences of the user. These systems work by matching the characteristics of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:27 UTC from IEEE Xplore. Restrictions apply.

the item to the customer's past behavior, which allows them to make accurate recommendations. CBRS can be further divided into two types: feature-based and model-based [2].

3) Hybrid Recommender System: It combines different types of recommenders to provide more accurate and personalized recommendations. These systems can be constructed in several ways, such as by combining CF and CBRS or using multiple CF algorithms with different weightings [3].
4) Graph-Based Recommender Systems (GBRS): It is based on graph theory and uses graph structures to model the interconnections between users and items. GBRS are particularly useful when dealing with sparse data sets, and they can be used in conjunction with other types of recommenders [4].
5) Community-Based Recommender System: It uses community detection algorithms to group customers and products into clusters based on similarity. These clusters can then be used to produce recommendations based on the behavior of similar users or items within the same cluster.

Different types of recommenders have different pros and cons, and the choice of the appropriate system depends on the nature of the problem and the available data.

B. Use of Community Detection to improve Recommender Sytems

Community detection, a popular method in network analysis, can help recommender systems to give more accurate recommendations by identifying communities or groups of users with similar preferences or behavior.

Community detection can be used to recognize clusters of customers with similar preferences or behavior, which can be used to increase the accuracy of recommender systems. For example, Gasparetti et al. [5] used community detection to enhance the accuracy of a social recommender system by identifying communities of users with similar interests. They found that incorporating community detection into the recommendation process improved the accuracy of the system, especially for users with sparse data.

Another way community detection can be used to enhance recommender systems is by identifying influential users or nodes in the network. Influential users or nodes are those that have a remarkable impact on the preferences or behavior of other users in the network. Identifying these users can help to improve the accuracy of recommendations by giving more weight to their preferences.

In addition to identifying communities and influential users, community detection can also be used to improve the diversity of recommendations. Diversity is an important factor in recommender systems, as users may want to explore new items or experiences. By identifying communities of users with different preferences or interests, recommender systems can provide more diverse recommendations. For example, Zhao et al. [6] used community detection to improve the diversity of recommendations in a social recommender system by identifying communities of users with different preferences and providing recommendations from each community.

Another way community detection can be used to enhance recommender systems is by incorporating temporal information into the analysis. Temporal information, for example, the time of user activity or the order in which items were consumed, can provide valuable insights into user behavior and preferences. Community detection can be used to identify temporal patterns in the network, which can be used to improve the accuracy of recommendations. For example, Chang et al. [6] used community detection to analyze temporal patterns in a network of users and items and found that incorporating temporal information improved the accuracy of a recommender system.

Community detection can also be used to improve the scalability and efficiency of recommender systems. Traditional recommender systems may struggle to handle large-scale networks with millions of users and items. Community detection can be used to reduce the complexity of the network by identifying communities of users with similar preferences or behavior. This can make the recommendation process more efficient and scalable. For example, the Louvain community detection algorithm [7] was proposed as a fast community unfolding method for large networks.

Overall, community detection is very important for improving the accuracy, diversity, and scalability of recommender systems. By clustering users based on their behavior within a network, community detection algorithms can identify latent user preferences, handle the cold-start problem, identify items that are popular within particular communities, improve the diversity of recommendations, and improve the scalability of recommender systems. These benefits have led to a significant increase in research on community detection for recommender systems in the last few years and are likely to continue driving innovation in the field in the future.

C. Comparison of traditional recommender systems with community-based recommender systems

Traditional recommender systems tend to use collaborative filtering and content-based approaches to generate recommendations. However, these approaches often suffer from the cold start problem, sparsity problem, and lack of diversity. In contrast, community-based recommender systems use community detection algorithms to identify groups of users with similar preferences and behaviors, leading to more accurate and diverse recommendations. Additionally, these systems can also incorporate social influence and homophily effects, further improving their recommendation quality.

[6] provides a comparison between traditional recommender systems and community-based recommender systems. The authors note that community-based approaches have shown to be more effective in handling the cold start problem by leveraging the social connections between users. Similarly, [5] emphasizes that community-based approaches can mitigate the sparsity problem by identifying groups of users

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:27 UTC from IEEE Xplore. Restrictions apply.

with similar preferences and increasing the overlap in their item ratings. The authors of [5] also argue that community-based recommender systems can achieve better performance than traditional methods by incorporating social influence and homophily effects, which are not captured by traditional collaborative filtering approaches.

Furthermore, several studies have compared the performance of community-based recommender systems with traditional recommender systems. For example, [8] compares the performance of community-based and traditional approaches on several benchmark datasets and shows that community-based approaches outperform coventional approaches in terms of accuracy and diversity. Similarly, [7] compares several community detection algorithms and shows that community-based approaches can achieve higher modularity and clustering quality than traditional collaborative filtering approaches.

In addition to accuracy and diversity, community-based recommender systems can also give better explanations for their recommendations. [5] argues that community-based approaches can give more transparent and interpretable recommendations by associating them with specific user communities.

Overall, community detection can be a powerful tool to improve the accuracy of recommender systems. Community-based recommender systems can address many challenges faced by traditional approaches, including the cold start problem, sparsity problem, and lack of diversity. Moreover, these systems can incorporate social influence and homophily effects, leading to more accurate and diverse recommendations. Additionally, community-based approaches can give better explanations for their recommendations, making them more transparent and interpretable to users.

III. Community Detection Algorithms for Recommender Systems

Community detection algorithms have proven to be effective tools for improving recommender systems. These algorithms aim to identify communities of users with similar preferences or behaviors, which can help to understand user interactions better and provide more accurate recommendations. In this section, we provide an overview of different community detection algorithms that can be applied to recommender systems, along with a description of how each algorithm works and its strengths and limitations [9].

One popular community detection algorithm for recommender systems is the Louvain algorithm [7]. The Louvain algorithm is an iterative algorithm that optimizes the network's modularity by iteratively moving nodes to different communities to maximize the modularity score. This algorithm is known for its speed and scalability, making it ideal for large networks. However, it can struggle with identifying communities with overlapping nodes, which is a common issue in recommender systems where users may have diverse preferences [10].

Another algorithm that has been successfully applied to recommender systems is the Label Propagation algorithm [6]. The Label Propagation algorithm works by propagating labels between nodes until the network is partitioned into communities. This algorithm is efficient and can identify overlapping communities, which makes it well-suited for recommender systems. However, it can struggle with detecting communities with sparse connections.

The Infomap algorithm is another community detection algorithm that has been used in recommender systems [8]. The Infomap algorithm works by partitioning the network into modules based on information flow. It is known for its ability to identify hierarchical community structures and its ability to detect overlapping communities. However, it can be computationally expensive and can struggle with detecting communities with few nodes.

The Spectral Clustering algorithm is another community detection algorithm that has been used in recommender systems [6]. Spectral Clustering works by identifying clusters based on the eigenvectors of the network's adjacency matrix. It is known for its ability to handle high-dimensional network data, making it well-suited for recommender systems with large amounts of data. However, it can struggle with identifying communities with overlapping nodes.

One final community detection algorithm that has been used in recommender systems is the Edge Betweenness algorithm [6]. The Edge Betweenness algorithm works by removing edges with high betweenness centrality, which helps to identify communities of nodes that are tightly connected. This algorithm is efficient and can handle networks with overlapping communities. However, it can struggle with detecting communities with a low density of connections.

Overall, each community detection algorithm has its own pros and cons, making it important to carefully choose an appropriate algorithm based on the specific characteristics of the recommender system being analyzed. For example, the Louvain algorithm may be the best choice for large-scale networks, while the Label Propagation algorithm may be best for identifying overlapping communities. By using these algorithms to recognize communities of users with similar preferences, recommender systems can better acknowledge user interactions and provide better recommendations.

IV. Applications of Community Detection in Recommender Systems

Community detection algorithms have been applied in a variety of ways to enhance the quality of recommendations and enhance user experience in recommender systems. One application is to use community detection to recognize groups of customers with similar preferences and recommend items that are favourite within their respective communities. This approach has been used successfully in various real-world recommender systems, such as Netflix, to improve the accuracy of recommendations.

Another application of community detection in recommender systems is to identify user communities based on their social connections or network structures. This approach has been used in social recommendation systems, such as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:27 UTC from IEEE Xplore. Restrictions apply.

Facebook and LinkedIn, to recommend friends or professional connections based on common interests or affiliations.

Community detection can also be used to solve the "cold-start" problem in recommender systems, where new customers or products have no or limited historical data for recommendations. By identifying communities of similar users or items, recommender systems can leverage the preferences and behaviors of existing users or items in those communities to make recommendations for new customers or products.

Another application of community detection is to improve the diversity of recommendations by identifying and recommending items from different communities or clusters. This can help overcome the problem of over-recommending popular or mainstream items, and provide users with a more diversified set of recommendations.

Community detection can also be used to enhance the explainability of recommendations. By identifying the communities or clusters of products that are recommended to a customer, recommender systems can provide users with more transparent and interpretable recommendations, explaining why certain products are recommended based on their similarity to items preferred by other users in the same community [11].

In addition, community detection can be used to enhance the serendipity of recommendations, by identifying unexpected or novel items that are preferred by users in a given community, and recommending them to users outside that community. This can help increase user satisfaction and engagement, as users may discover new items that they would not have otherwise found on their own.

There are several real-world examples of successful implementations of community detection in recommender systems. For instance, Amazon uses community detection algorithms to group products and generate personalized recommendations for individual users. Facebook uses community detection algorithms to identify groups of customers with similar interests and generate personalized news feeds. LinkedIn uses community detection algorithms to identify groups of customers with similar professional backgrounds and generate personalized job recommendations.

Overall, community detection has been shown to be a valuable tool in improving the quality and user experience of recommender systems. By leveraging the similarities and connections between users and items, community detection algorithms can identify meaningful and relevant communities, and make recommendations that are more accurate, diverse, explainable, and serendipitous.

V. Evaluation Metrics for Community-based Recommender Systems

Community based recommender systems (CBRS) have gained significant attention in recent years due to their ability to provide personalized and accurate recommendations by leveraging community structure in the data. However, evaluating the performance of CBRS presents a unique challenge, as traditional evaluation metrics used for recommender systems may not be applicable or appropriate for community-based approaches. Therefore, researchers have proposed and adapted various evaluation metrics to calculate the performance of CBRS.

One commonly used metric for evaluating the performance of CBRS is Normalized Discounted Cumulative Gain (NDCG) [12], which is a widely used metric in traditional recommender systems. NDCG measures the ranking quality of the recommended items, taking into account both relevance and position of the items in the ranked list. Another commonly used metric is Precision, which measures the fraction of relevant items in the top-K recommended items. Both NDCG and Precision have been adapted for community-based approaches and widely used to evaluate their performance.

Community-based Precision (CP) is another metric that has been proposed specifically for community-based approaches [13]. CP measures the proportion of items recommended from the same community as the user, indicating the ability of the CBRS to capture community preferences. A related metric is Intra-Community Precision (ICP), which measures the proportion of recommended items from the same community as the user that are also relevant to the user. These metrics reflect the ability of the CBRS to capture both the user's individual preferences and the preferences of their community.

Other evaluation metrics that have been proposed for CBRS include Diversity and Coverage. Diversity measures the variety of recommended items, and Coverage measures the proportion of items that are recommended at least once. These metrics are important in ensuring that the CBRS provides an extensive range of recommendations that appeal to a wide range of users.

When comparing evaluation metrics between CBRS and traditional recommender systems, it is important to note that CBRS typically outperform traditional approaches in terms of NDCG and Precision. However, traditional approaches tend to perform better in terms of Diversity and Coverage. This suggests that CBRS are better suited for providing personalized recommendations to users within specific communities, while traditional approaches are better suited for providing a broad range of recommendations to a wider audience.

In conclusion, evaluating the performance of CBRS presents unique challenges, and traditional evaluation metrics may not be applicable or appropriate. Metrics such as NDCG, Precision, CP, ICP, Diversity, and Coverage have been proposed specifically for CBRS and have been shown to be effective in evaluating their performance. When comparing evaluation metrics between CBRS and traditional recommender systems, it is essential to consider the specific goals and audiences of each approach.

VI. Challenges

Data sparsity is a common problem in recommender systems, where most users only rate a small fraction of the available items. This results in a sparse user-item rating matrix that is difficult to analyze and model. When applying community detection to such data, the sparsity problem becomes even more pronounced, as the number of edges between customers

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:27 UTC from IEEE Xplore. Restrictions apply.

and products decreases, making it difficult to detect meaningful communities. To overcome this challenge, researchers have proposed various techniques, such as incorporating external data sources and leveraging user social networks to supplement the user item rating matrix [14].

Scalability is another significant challenge faced in applying community detection to recommender systems. As the size of the user-item rating matrix increases, the computational complexity of community detection algorithms also increases, making it difficult to perform community detection in real-time on large-scale datasets. To address this challenge, researchers have proposed several scalable community detection algorithms, such as MapReduce-based algorithms and distributed clustering algorithms.

Another challenge faced in applying community detection to recommender systems is the interpretation of the detected communities. In many cases, the detected communities may not have a clear interpretation in terms of user preferences or interests. This makes it difficult to use the detected communities to improve the quality of recommendations. To address this challenge, researchers have proposed various methods, such as incorporating domain knowledge and incorporating explicit feedback from users.

In addition to the challenges discussed above, applying community detection to recommender systems also raises privacy concerns, as the analysis of user-item ratings may reveal sensitive information about users' preferences and behaviors. Therefore, it is essential to ensure that appropriate privacy-preserving techniques are employed when applying community detection to recommender systems [15].

In conclusion, applying community detection to recommender systems presents several challenges, including data sparsity, scalability, and interpretation of the detected communities. These challenges require innovative solutions, such as incorporating external data sources, developing scalable algorithms, incorporating domain knowledge, and employing privacy-preserving techniques. Addressing these challenges is critical for improving the accuracy of recommendations in recommender systems and providing personalized experiences to users.

VII. Future directions for research

Community detection in recommender systems has emerged as a promising technique for personalized and accurate recommendations. However, there is still a lot of research to be done in this area, especially in incorporating user feedback and social network information.

One future direction for research is the integration of user feedback into the community detection process. Currently, community detection algorithms are mostly unsupervised and rely on the underlying data structure to identify communities. However, user feedback, such as ratings or reviews, can provide valuable information about the user's preferences and interests, which can be used to refine the community structure. Several studies have proposed incorporating user feedback into the community detection process, such as the work by [16], who proposed a feedback-based community detection algorithm that incorporates user preferences to improve the accuracy of the recommendations.

Another area for future research is the integration of social network information into the community detection process. Social networks can provide valuable information about the relationships between users and the communities they belong to, which can be used to enhance the accuracy and relevance of the recommendations. Several studies have explored the use of social network information in community detection for recommender systems, such as the work by [17], who proposed a social network-based community detection algorithm that consider the relationships between users and their social network.

Furthermore, the scalability of community detection algorithms in the context of recommender systems is another area for future research. With the increasing size of data in recommender systems, traditional community detection algorithms may not be scalable and efficient enough to handle large datasets. Several studies have proposed scalable community detection algorithms, such as the work by [18], who proposed a fast algorithm for detecting communities in large-scale networks based on the MapReduce framework.

In addition to the above, there is also a need to explore the use of deep learning for community detection in recommender systems. Deep learning techniques, such as neural networks, have shown promising results in various fields and can potentially enhance the accuracy and scalability of community detection algorithms in recommender systems. Several studies have explored the use of deep learning for recommender systems, such as the work by [19], who proposed a deep neural network based model for personalized recommendation.

In conclusion, there are several future directions for research in community detection for recommender systems, including the incorporation of user feedback and social network information, scalability of algorithms, and the exploration of deep learning techniques. These directions can potentially enhance the accuracy, relevance, and scalability of community-based recommender systems and provide personalized recommendations to users.

VIII. Case Studies and Use Cases

Community detection in recommender systems has shown great potential to enhance the performance of recommendation algorithms in various industries such as social media, e-commerce, and entertainment. In this section, we will discuss several case studies and use cases that illustrate the effectiveness of community detection in recommender systems.

One example of the successful implementation of community detection technique in recommender systems is Amazon's "Customers Who Bought This Item Also Bought" feature. Amazon uses community detection algorithms to recognize groups of users with similar purchase histories, and then recommends items based on the purchasing behavior of those groups. This approach has been shown to increase customer satisfaction and drive sales [20].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:27 UTC from IEEE Xplore. Restrictions apply.

Another example is the social media platform Twitter, which uses community detection algorithms to suggest relevant accounts for users to follow based on their interests and activity. Twitter's "Who to Follow" feature utilizes community detection to identify users who share similar interests and recommends them to each other [21].

In the entertainment industry, community detection has been used to improve movie recommendations on platforms such as Netflix. Netflix's recommendation algorithm utilizes community detection to identify groups of users with similar viewing histories and recommends movies based on the preferences of those groups. This approach has been shown to significantly increase user engagement and satisfaction with the platform.

Community detection has also been applied in the healthcare industry to improve patient outcomes. For example, community detection algorithms have been used to identify groups of patients with similar medical histories and recommend personalized treatment plans based on the preferences of those groups.

In addition to the industries mentioned above, community detection has also been successfully applied in other domains such as education, finance, and transportation. For example, in education, community detection algorithms are used to identify groups of students with similar learning preferences and recommend personalized learning materials [22]. In finance, community detection algorithms are used to identify groups of investors with similar investment portfolios and recommend personalized investment strategies. In transportation, community detection algorithms are used to identify groups of commuters with similar travel patterns and recommend personalized commuting routes.

In conclusion, community detection has shown great potential to enhance the performance of recommender systems in various industries. The case studies and use cases discussed above demonstrate the effectiveness of community detection in providing personalized and relevant recommendations to users. As technology continues to advance, it is anticipated that community detection algorithms will become increasingly sophisticated and will continue to be applied in new and innovative ways.

IX. Conclusion

Recommender systems are extensively used for information retrieval. Community detection techniques are used to enhance the performance of recommender systems. In this paper, we have presented a review of various types of recommender systems and the various applications and algorithms of community detection for the recommender systems. Further, we discussed the challenges in using community detection for recommender systems and concluded with the research directions and case studies of community detection based recommender systems.

References

[1] Schafer, J. Ben, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative filtering recommender systems." The adaptive web: methods and strategies of web personalization (2007): 291-324.

[2] Aggarwal, Charu C., and Charu C. Aggarwal. "Content-based recommender systems." Recommender systems: The textbook (2016): 139-166.

[3] Çano, Erion, and Maurizio Morisio. "Hybrid recommender systems: A systematic literature review." Intelligent Data Analysis 21, no. 6 (2017): 1487-1524.

[4] Chicaiza, Janneth, and Priscila Valdiviezo-Diaz. "A comprehensive survey of knowledge graph-based recommender systems: Technologies, development, and contributions." Information 12, no. 6 (2021): 232.

[5] Gasparetti, Fabio, Giuseppe Sansonetti, and Alessandro Micarelli. "Community detection in social recommender systems: a survey." Applied Intelligence 51 (2021): 3975-3995.

[6] Su, Xing, Shan Xue, Fanzhen Liu, Jia Wu, Jian Yang, Chuan Zhou, Wenbin Hu et al. "A comprehensive survey on community detection with deep learning." IEEE Transactions on Neural Networks and Learning Systems (2022).

[7] https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae

[8] Souravlas, Stavros, Angelo Sifaleras, M. Tsintogianni, and Stefanos Katsavounis. "A classification of community detection methods in social networks: a survey." International Journal of General Systems 50, no. 1 (2021): 63-91.

[9] Kumar, Sanjay, and Rahul Hanot. "Community detection algorithms in complex networks: A survey." In Advances in Signal Processing and Intelligent Recognition Systems: 6th International Symposium, SIRS 2020, Chennai, India, October 14–17, 2020, Revised Selected Papers 6, pp. 202-215. Springer Singapore, 2021.

[10] Kumar, Sanjay, B. S. Panda, and Deepanshu Aggarwal. "Community detection in complex networks using network embedding and gravitational search algorithm." Journal of Intelligent Information Systems 57 (2021): 51-72.

[11] Zhang, Yongfeng, and Xu Chen. "Explainable recommendation: A survey and new perspectives." Foundations and Trends® in Information Retrieval 14, no. 1 (2020): 1-101.

[12] Järvelin, Kalervo, and Jaana Kekäläinen. "Cumulated gain-based evaluation of IR techniques." ACM Transactions on Information Systems (TOIS) 20, no. 4 (2002): 422-446.

[13] Feng, Liang, Qianchuan Zhao, and Cangqi Zhou. "Improving performances of Top-N recommendations with co-clustering method." Expert Systems with Applications 143 (2020): 113078.

[14] Gasparetti, Fabio, Alessandro Micarelli, and Giuseppe Sansonetti. "Community Detection and Recommender Systems." (2018).

[15] Fung, Benjamin CM, Ke Wang, Rui Chen, and Philip S. Yu. "Privacy-preserving data publishing: A survey of recent developments." ACM Computing Surveys (Csur) 42, no. 4 (2010): 1-53.

[16] Tang, Lei, and Huan Liu. "Community detection and mining in social media." Synthesis lectures on data mining and knowledge discovery 2, no. 1 (2010): 1-137.

[17] Ma, Hao, Haixuan Yang, Michael R. Lyu, and Irwin King. "Sorec: social recommendation using probabilistic matrix factorization." In Proceedings of the 17th ACM conference on Information and knowledge management, pp. 931-940. 2008.

[18] Leskovec, Jure, Kevin J. Lang, and Michael Mahoney. "Empirical comparison of algorithms for network community detection." In Proceedings of the 19th international conference on World wide web, pp. 631-640. 2010.

[19] Zhang, Shuai, Lina Yao, Aixin Sun, and Yi Tay. "Deep learning based recommender system: A survey and new perspectives." ACM computing surveys (CSUR) 52, no. 1 (2019): 1-38.

[20] Linden, Greg, Brent Smith, and Jeremy York. "Amazon. com recommendations: Item-to-item collaborative filtering." IEEE Internet computing 7, no. 1 (2003): 76-80.

[21] Huberman, Bernardo A., Daniel M. Romero, and Fang Wu. "Social networks that matter: Twitter under the microscope." arXiv preprint arXiv:0812.1045 (2008).

[22] Koper, Rob, and Bill Olivier. "Representing the learning design of units of learning." Journal of Educational Technology & Society 7, no. 3 (2004): 97-111.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:27 UTC from IEEE Xplore. Restrictions apply.

2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE) | 979-8-3503-1312-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICEPE57949.2023.10201637

A Review on Recent Encoder-Decoder Architectures for Drivable Region Detection and Segmentation

Dhruv Sharma
Department of Electronics and Communication
Delhi Technological University
New Delhi, India
dhrv99@gmail.com

S. Indu
Department of Electronics and Communication
Delhi Technological University
New Delhi, India
s.indu@dce.ac.in

N. Jayanthi
Department of Electronics and Communication
Delhi Technological University
New Delhi, India
njayanthi@dce.ac.in

***Abstract*— With the proliferation of science and technology, autonomous driving has garnered significant attention in the recent years due to the manifold benefits of driverless vehicles, including enhanced quality of life, mitigation of accidents, and curtailment of traffic congestion. In this work, recently developed encoder-decoder oriented architectures specifically designed to detect and segment the safe drivable region on roads have been reviewed, analyzed and compared, and the challenges in this field have been discussed. In pursuit of this objective, we initially furnish a survey of various orthodox and deep learning methods that find their use in detection of the road obstacles, and segmentation of roads and lane-lines. This is followed by a discussion on relevant field-oriented datasets. Subsequently, we present some recent encoder-decoder based methodologies such as HybridNets, You Only Look Once Panoptic (YOLOP), etc. These methods are then compared in terms of both inference speeds as well as performance metrics, to understand their potential in the real-time better, concluding that YOLOPv2 is currently the best suited architecture meant for drivable region detection. This is followed by an examination of challenges and unresolved inquiries in this domain.**

Keywords—autonomous driving, detection, encoder-decoder architecture, segmentation.

I. INTRODUCTION

With an expedited breakthrough of the transportation sector over the past century, humans now manifest a radical transformation in this domain leading to an eager interest in the field of driverless transportation. A driverless vehicle is the one that is utterly self-operated and requires no intervention from the human. Synonymously known as "autonomous" or "self-driving" vehicles, they incorporate software and sensor technologies together in order to navigate and propel a vehicle. Given that human error is the key contributory factor causing road accidents [1], driverless technology holds great potential for reducing the number of such incidents on roads [2]. Hence, apart from improvement in the lifestyle, driverless vehicle technology also makes the vehicles less prone to accidents. Such systems are meritorious in bringing down the logjam and traffic congestion.

Detection of the drivable-region forms a crucial chunk of driverless vehicle operations. Accurately detecting the drivable area is crucial for the safe and efficient operation of autonomous vehicles. Drivable area detection has been successfully tackled using deep learning, which has emerged as a powerful solution for complex computer vision problems. In this regard, deep learning models are utilized to assess the safe and drivable portion of the road for a vehicle upon being trained with large sets of labeled data to identify distinguishing aspects and patterns that warrant the distinction between drivable and non-drivable regions. The propitious results of this approach have shown its potential in nurturing the automated driving systems. A typical detection system for drivable region consists of obstacle detection (vehicles, animals, humans, etc.), drivable region detection, lane detection, pothole detection, traffic signs detection, etc [3].

Considerable progress and research have been made in this domain over the last couple of decades. The archaic methods including geometry-based modelling and energy- oriented optimization schemes have evolved into newer clutter-breaking deep learning-based methodologies. The geometry-oriented-modelling approach involves detecting the edges of the driveway and fitting lines over them. Several filtering-based techniques, such as Gaussian [4] and Gabor filtering [5] have been employed for edge detection [6], with a focus on gradients. Additionally, aspects such as texture and colour can also be incorporated. The Hough transform methods have been used to adjust the lines over the detected edges [7]. Energy optimization approaches, such as Continuous Random Fields (CRF), have also been utilized to detect lanes [8]. With the advent of deep learning these models were ousted. Such peculiar methods included object detectors useful to predict obstacles on the driveway [9]. Deep learning algorithms, such as Convolutional Neural Networks (CNNs) [10], have shown great potential in accurately identifying objects, how- ever at a greater inference time. Region-Convolution Neural Network (R-CNN) has its output in the form of obstacles' bounding boxes sets along with the class output [11]. Their improvements fast R-CNN [12] and faster R-CNN [13] were also instigated. However, owing to the large inference times during testing they do not cater to the real time execution. Single shot detector (SSD) provides good frame rate during inference wearing from 22-59 frames per second (FPS). They utilize single progressive pass to perform the steps of localization and categorization. However, their performance worsens for tinier objects [14]. You Only Look Once (YOLO) considers detecting of objects a regression problem it utilizes the technique of division of ground truth into grids hence only one forward propagation is the requisite to implement the methodology it offers good frame rate of 45 frames per second [15]. YOLO was further evolved into YOLOv2 [16] and v3 [17], as more effective object detectors. Methods of such kind have proved to be helpful in detecting traffic objects on roads.

Further several segmentation methodologies have also been put into use to detect the road [18] and lane-lines [19]. U-Net [20] classifies as each pixel of the input, henceforth,

979-8-3503-1312-3/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:48:32 UTC from IEEE Xplore. Restrictions apply.

being highly capable of differentiating borders of an object. It has contractive and expansive parts as the input and output are both supposed to be of the same dimensions. Unlike UNet that performs semantic segmentation, LaneNet [21] performs instance segmentation. It involves binary segmentation along with clustering-oriented embedding for cast graphic transformation. Spatial-CNN (SCNN) is capable of capturing translational and rotational relationships of frame pixels which apparently improve the segmentation process by manifolds [22]. In this model, rows and columns of feature maps are treated as layers with sequential passage of pixel details. Quin et al. [23] proposed a network that uses a hybrid architecture infusing CNN and Long Short-Term Memory (LSTM) [24]. CNN takes in data following a time series property while the LSTM learns the attribute and this is followed by Lane forecasting. In order to detect the entire driveway area, which would involve both detection of obstacles as well as segmentation of the lane-lines and the driveways, a combination of such networks is a must. However, opting for such a scenario would correspond to passing an input via three different models. This would reduce the speed and increase latency unreasonably and such models would fail to function well in real time scenarios. Hence, some researchers amalgamated such networks into much simpler architectures. Such architectures use simple encoder-decoder- based networks [25].

Mask R-CNN is an encoder-decoder-oriented network [26] that is capable of performing segmentation of images along with the detection of the objects. It is an extension of Faster- RCNN. While the former three versions of YOLO were introduced with an intent to detect objects, the latter versions including the YOLOv4 [27] and YOLOv7 [28] are extended to segmentation as well, along with the object detection. YOLOP is one such architecture that is capable of handling the three tasks specifically designed for driveway area detection [29]. It consists of an encoder that performs abstraction of the attributes. The encoder consists of a backbone and a neck. For backbone the authors have used CSP dark net while for the next spatial pyramid pooling has been used. It also has three decoders that performed three corresponding tasks. The model offers a speed of 41 fps. On the other hand, the DLT-Net is another architecture that is used to detect the driveway areas the lane and the traffic-oriented objects. Unlike in the other two architectures, it uses context tensors in order to amalgamate the decoders to perform the said tasks. It has an encoder based on the VGG-16 network [30]. HybridNets constitute of a common encoder and two decoders for separate vehicle detection and lane-driveway area segmentation [31]. The encoder uses a neck and backbone. The former is based on EfficientDet. Backbone of the encoder based on EfficientNet-B3. Bidirectional-Feature Pyramidal-Network (BiFPN) has been used for model scaling. The idea is to reduce the latency and increase the inference speeds while detecting without compromising much with the performance metrics, hence finding an optimal trade-off between the two significant requirements in order to make these models competent enough to function efficiently in the real world. YOLOv2 [32] has been developed, drawing inspiration from YOLOP and HybridNets, with YOLOv7 serving as the base model. However, even these methods have limitations such as exclusion of road damages and potholes detection mechanisms from such models could make them less guarded against accidents.

In this paper, we have reviewed some of the commonly used road related datasets and the latest encoder-decoder architectures, specifically designed for the application of drivable-region detection and segmentation. To the utmost extent of our cognizance and understanding, the previous works haven't primarily focused on analyzing encoder-decoder architectures specifically designed for this application, even though they are extremely meritorious. Additionally, we have also laid emphasis on both performance in terms of accurate detection as well as their real-time feasibility, which helps in understanding their potential and impact to a deeper extent. We have also analyzed and compared these architectures in terms of speed and evaluation metrics, and analyzed their limitations and challenges.

The paper is structured as follows: Section II discusses the commonly used datasets that are based on driving scenarios. Section III covers the general methodology and recent encoder-decoder detection methods for drivable-region detection. This is followed by comparison between these methods in section IV, followed by challenges in the field in Section V and section VI summarizes this study.

II. Dataset

A numerous road-based datasets have been introduced in the recent times. KITTI dataset [33] comprises of 15k road scene images. Using eight distinct classes, the dataset aims at covering applications including detection, segmentation and depth recovery. The dataset covers over 200k annotations corresponding to the object detection. It also intends to minimize overfitting in the scenarios. There's an inclusion of diverse road scenarios in Cityscapes dataset [34]. It consists of 25k road images covering high-quality as well as coarse annotations. The main idea behind this dataset was to cover more in-depth scenarios that are capable of replicating real life road conditions better. However, this dataset doesn't cover rainy and night road scenes like KITTI. The Oxford RoboCar dataset [35] primarily concentrated on autonomy in terms of its longevity. They covered the same routes under distinct conditions with an intent to capture different appearances of term predictions. nuScenes dataset [36] majorly covers

TABLE I. Dataset Comparison

Dataset	Year	Images	Classes	Rainy/Night scenes	Applications
KITTI [33]	2012	15k	8	No/No	Detection, Segmentation and Depth recovery
Cityscape [34]	2016	25k	30	No/No	Instance/Semantic Segmentation
Oxford RoboCar [35]	2017	64.8M	-	Yes/Yes	Pose prediction, long-term mappingand Localization
nuScenes [36]	2019	1.4M	23	Yes/Yes	Detection, Segmentation and Tracking
BLVD [37]	2019	120k	3	-- /Yes	Annotations, Trajectory Projections and Event recognition
BDD100k [38]	2020	100M	10	Yes/Yes	Detection, Segmentation, Tracking and Imitation learning

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:48:32 UTC from IEEE Xplore. Restrictions apply.

obstacle detection, road segmentation as well as object tracking applications. It is also focused upon pedestrian posing and vehicles' movement related applications. It contributes towards improved scene understanding by incorporating a 360-degree coverage of the road scenes. Build a Large-scale 5D Semantics Benchmark (BLVD) is another dataset [37] containing 120K road scene images. These images vary from lower to higher density scenarios on roads in terms of congestion and cover different times of a day. In addition to the aforementioned annotation aspects, BLVD also focuses on advanced higher dimensional applications including trajectory prediction and 5-D event identification that comprises of events like vehicle overtaking, pedestrian crossing and riding closer/farther, to name a few. BDD100K dataset [38] comprises of 100k road videos that are heavily inspired from real-life distinct and evaluation of ten different tasks including Multiple Object Tracking (MOT), MOT with segmentation, scene tagging, detection, imitation learning, etc. The datasets have been summarized in table I.

III. Methods

The general methodology for safe drivable area detection is summarized in the flowchart shown in figure 1. Images covering driving scenarios are first annotated using annotation tools into appropriate formats. There are various annotation formats including the COCO, Pascal VOC and YOLO [39]. The annotations, along with the ground truth (Images) are fed as inputs to the deep-learning architecture. It is an encoder-decoder based architecture. The encoder part consists of the backbone and the neck. It is used to extract features in the form of feature maps from the input training data. This is followed by multiple decoders used to effectuate a variety segmentation and detection processes to obtain the training and validation results. The optimal model is selected as the trained model, to which the testing data from the real-world is fed, and finally, the segmentation and detection results of the real-world data are obtained.

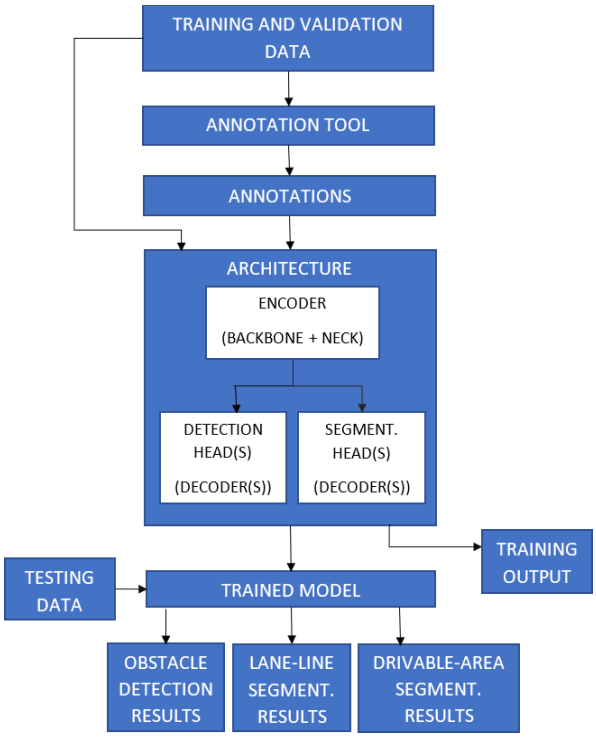


Fig. 1. General Block Diagram for Encoder-Decoder Architecture

A. DLT-Net

Drivable-area, Lane line and Traffic detection-Network (DLT-Net), indicating the three tasks involved in detecting drivable-region, is primarily focused upon the dependency of the three decoder tasks possible with the aid of the context provided by the specially designed tensors [30]. These tensors, henceforth, help in fusing feature maps of the three tasks. The encoder is based on VGG16 architecture and the pyramid structure. It utilizes the element mean operation to amalgamate the shallow and profound characteristics. The encoder generates feature maps of five varying sizes, of which, the concluding feature map is fed to the three task-oriented decoders. The context tensors, used in between the encoder and the decoders, provide aid in improving the discriminatory factor between the three tasks. The traffic detection decoder network is divided into two parts, one meant to perform classification and other for the bounding-box regression. It has several shortcomings. DLT-Net is incapable of giving optimal results during severe reflection driving scenarios. Also, it is incompetent in predicting the recurring divider lines as lane-lines.

B. DFM-RTFNet

Dynamic Fusion Module-RGB Fusion Thermal Network (DFM-RTFNet) is based on feature fusion technique, done dynamically by infusing DFM module to the RTFNet model [40]. It generates a dynamic kernel based on a two-dimensional feature and applies it on another feature to produce the fused feature. Owing to their properties including spatial variance and dependency on the content, optimum distinct kernels can be applied to different regions of an image. This in turn helps in optimizing the fusion process, thereby improving the efficiency of the network. To avoid excessive usage of the memory, the authors also constructed a dual-stage factorization-oriented convolution method. These DFMs then dynamically fuse two different modalities of features in a multiscale fashion. The RTF-Net uses two encoders for RGB and Thermal inputs. These encoders are inspired from the ResNet. Thereafter, five layers of decoders are utilized in order to process the fused features. Sequentially, this is followed by a softmax layer post which, the concluding detection and segmentation results are produced.

C. YOLOP

YOLOP is among first deep learning models for driveway-area detection that was realized by the authors on the Jeston Tx2, an embedded device to ensure its usability in real-time scenarios. The model was introduced with an intent to prosper in traffic object-detection, drivable-region segmentation and lane-line detection. It is primarily developed from its base model YOLOv4 with Adam being used as the optimizer for training [29]. The encoder consists of backbone and neck architectures with the former based on Cross Stage Partial Dark Network (CSPDarkNet) and the latter on Spatial Pyramid Pooling (SPP) module. It is based on scaled YOLOv4 which is quintessentially a lightweight CNN. The detector part comprises of three separate decoders. Decoder performs anchor-oriented multiscale detection and uses the single-stage YOLOv4 model, owing to the model's excellent inference speed. It uses an amalgamation of Feature Pyramid Network (FPN) and Aggregation Network (PAN) (or bottom- up FPN) for the transference of semantic aspects top-down and positioning aspects bottom-up, respectively in order to obtain superior feature fusion effect. The bottom layer of the neck serves as the input to these heads. These heads involve triple-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:48:32 UTC from IEEE Xplore. Restrictions apply.

upsampling for the restoration of output feature map to the relevant size. This method, inspired from the trainable Bag of Freebies (BoF) helps in acquiring high-precision throughput and swifter inference.

D. HybridNets

Unlike in YOLOP and other models, HybridNets constitutes of a common encoder and only two decoders for separate vehicle detection and lane-driveway area segmentation, instead of three decoders [31]. Encoder has been designed to infuse model scaling which is extremely crucial for its hardware system implementation. Bidirectional FPN has been used for model scaling. Numerous loss functions are converged with the aid of a special learning scheme based on multistage methods. EfficientNet-B3 serves as the backbone. Its major role involves optimization of network by evaluating scaling parameters. This makes the network design stable. The encoder uses neck network based on EfficientDet which receives the feature maps from the backbone. BiFPN combines attribute at a varied resolution. This is based upon the cross-scale connection for each node by each bidirectional path. It also builds on weight corresponding to each attribute in order to comprehend the significance of each level. K-means clustering is utilized to establish the anchor boxes that're assigned to fusion maps. The model is capable of working well on complicated datasets. The background, driveway-area and lane-line are the three categories to be examined by the segmentation head.

E. YOLOPv2

YOLOPv2 was designed, drawing inspiration from existing works such as YOLOP and HybridNets, but with modifications for improved efficiency. They also utilize a shared encoder and three separate decoder heads for specific tasks of drivable area segmentation and lane detection. The architecture is primarily based on YOLOv7 [32]. Group convolution from E-ELAN is employed in the backbone for diverse feature learning, and feature fusion is achieved through SPP and FPN modules. An anchor-based multi-scale detection scheme similar to YOLOv7 is adopted, utilizing PAN for localization feature extraction. However, drivable area segmentation and lane segmentation are performed in separate task heads with different network structures, as features from deeper layers are found to be unnecessary and even detrimental for drivable area segmentation. For segmentation, semantic-level features are fed as inputs. An additional upsampling layer is applied to compensate for this change, while deconvolution is used in the decoder stage of lane detection for improved performance. The model utilizes cosine annealing for learning-rate calibration in order to improve its performance. YOLOPv2 is also based on trainable BoF, hence making it more viable to solve real-world problems.

IV. COMPARISON

We compare the above-described models and some other state-of-the-art methods for detection of the safe drivable-region in different road-scenarios. The most commonly used performance metrics for obstacle detection is mean Average Precision (mAP). For drivable-region segmentation, Intersection over Union (IoU) metric is widely used, while for lane-line segmentation, accuracy is preferred over IoU since lane-lines are usually discontinuous and of variable widths. Performance metrics for the three tasks are summarized in Table II. PSP-Net, utilized on the Cityscapes, [41] uses pyramid pooling module and an FCN-oriented network to segment the roads and obstacles. It doesn't perform lane-line segmentation. The remaining techniques are capable of performing two or more detection/segmentation tasks useful in identifying the safe drivable-region. While E-Net, Multi-Net and DFM-RTFNet algorithms were originally applied on KITTI and Cityscapes datasets, but in [30] MultiNet was retrained on BDD100k dataset, so we used the same to comparison easier with the dataset of most of the major techniques common as the more recent methods have been applied on the BDD100k dataset. From the table, several conclusions maybe drawn. It may be concluded that YOLOPv2 is overall a more superior method performing all three tasks. It offers the most superior results for lane-line and drivable-region segmentation, offering an accuracy of 87.31% and IoU of 93.2%, respectively on BDD100k dataset. This is closely followed by HybridNets with its accuracy and IoU being 85.4% and 90.5 % respectively, and YOLOP with 70.5% and 91.5%, respectively. Even though YOLOPv2 also offers best mAP for traffic obstacle detection for BDD100k dataset, it is the dynamic aspect-fusion oriented DFM-RTFNet that offers the maximum mAP of all the models, applied on the KITTI dataset. However, it rules out the possibility of segmentation of the lane-lines, restricting it to perform only two tasks. In terms of speed, ENet outperforms the other models, operating at the speed 100 Frames Per Second (FPS) [42], followed by YOLOPv2 that operates at the speed of 93 FPS. However, E-Net offers a poor accuracy for lane-line segmentation, hence, incapable of finding the trade-off between the speed and performance- metrics. YOLOPv2, on the other hand, owing to its trainable bag of freebies, offers the best possible trade-off between the speed and performance metrics of the three tasks, proving to be the most superior method in the recent times, closely followed by HybridNets and YOLOP.

TABLE II. COMPARISON BETWEEN LATEST METHODS

Method	Year	Dataset	Speed (FPS)	Obstacle Detection (mAP)	Lane-Line Accuracy Segmentation (%)	Drivable Area Segmentation (IoU)
PSPNet [41]	2017	Cityscapes	11.1	-	-	89.6
ERFNet [43]	2017	Cityscapes	22.8	-	-	68.7
ENet [42]	2016	Cityscapes	100	-	34.2	-
Multi-Net [44] [30]	2018	BDD100k	8.6	60.2	-	71.6
DLT-Net [30]	2019	BDD100k	9.3	68.4	-	72.1
DFM-RTFNet [40]	2021	KITTI	-	94.05	-	86.5
YOLOP [29]	2022	BDD100k	41	76.5	70.5	91.5
HybridNets [31]	2022	BDD100k	-	77.3	85.4	90.5
YOLOPv2 [32]	2022	BDD100k	93	83.4	87.31	93.2

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:48:32 UTC from IEEE Xplore. Restrictions apply.

V. Challenges

An efficient data labeling prior to model training is desirable in order to produce optimum detection and segmentation results. Manual annotators have been the key source for data labeling. However, the process would become tedious and time consuming in order to ensure diversity within a dataset that depicts a wide variety of road scenarios. Further, human errors can also arise in the process. Several labeling methods have been developed in order to make the data labeling less tedious [45]. There is still a scope of improvement in this field.

Latency is an important parameter and can't be compromised with especially in the real-time scenarios where a car is being driven at a high speed. If Latency is improved the performance metrics tend to decline substantially and vice-versa. Object detection and segmentation techniques individually offer superior performance metrics but their amalgamation will cause a decline in the model speed in terms of fps, hence miserably affecting the latency. Encoder-Decoder architectures provide improved latency with a significant decline in the performance metrics. To establish an efficient deep learning model that finds a trade-off between latency and metrics is one of the key challenges in this field. Even though methods like YOLOP and YOLOPv2 have addressed the issue by proposing for trainable BoF as the solution, there's still a scope of improvement in this aspect.

Detection of potholes on the roads to avoid uneasy travel and damage to the vehicles using such deep-learning methods is another major challenge in the drivable-region detection algorithms and systems. It's relatively difficult to differentiate such concave-sculpted depressions from the roads especially during night scenarios and poor weather conditions. There is still a lack of fusion systems for drivable-region detection that also infuse detection of potholes present on the roads.

VI. Conclusion

With the proliferation of science and technology, autonomous driving has garnered significant attention in the recent years due to the manifold benefits of driverless vehicles, including enhanced quality of life, mitigation of accidents, and curtailment of traffic congestion. In this work, recently developed encoder-decoder oriented architectures specifically designed to detect and segment the safe drivable region on roads have been reviewed, analyzed and compared, and the challenges in this field have been discussed. In pursuit of this objective, we initially furnish a survey of various orthodox and deep learning methods that find their use in detection of the road obstacles, and segmentation of roads and lane-lines. This is followed by a discussion on relevant field-oriented datasets. Subsequently, we present some recent encoder-decoder based methodologies and their comparison, followed by an examination of challenges and unresolved inquiries in the field.

References

[1] R. Chai, A. Tsourdos, A. Savvaris, S. Chai, Y. Xia, and C. P. Chen, "Multiobjective overtaking maneuver planning for autonomous ground vehicles," IEEE transactions on cybernetics, vol. 51, no. 8, pp. 4035– 4049, 2020.

[2] T. Liu, Q. hai Liao, L. Gan, F. Ma, J. Cheng, X. Xie, Z. Wang, Y. Chen, Y. Zhu, S. Zhang, et al., "The role of the hercules autonomous vehicle during the covid-19 pandemic: An autonomous logistic vehicle for contactless goods transportation," IEEE Robotics & Automation Magazine, vol. 28, no. 1, pp. 48–58, 2021.

[3] Y. Sun, W. Zuo, H. Huang, P. Cai, and M. Liu, "Pointmoseg: Sparse tensor-based end-to-end moving-obstacle segmentation in 3-d lidar point clouds for autonomous driving," IEEE Robotics and Automation Letters,vol. 6, no. 2, pp. 510–517, 2020.

[4] M. Aly, "Real time detection of lane markers in urban streets," in 2008 IEEE intelligent vehicles symposium, pp. 7–12, IEEE, 2008.

[5] J. C. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation," IEEE transactions on intelligent transportation systems, vol. 7, no. 1, pp. 20– 37, 2006.

[6] S. Zhou, Y. Jiang, J. Xi, J. Gong, G. Xiong, and H. Chen, "A novel lane detection based on geometrical model and gabor filter," in 2010 IEEE Intelligent Vehicles Symposium, pp. 59–64, IEEE, 2010.

[7] A. Borkar, M. Hayes, and M. T. Smith, "Polar randomized hough transform for lane detection using loose constraints of parallel lines," in 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1037–1040, IEEE, 2011.

[8] J. Hur, S.-N. Kang, and S.-W. Seo, "Multi-lane detection in urban driving environments using conditional random fields," in 2013 IEEE IntelligentVehicles Symposium (IV), pp. 1297–1302, IEEE, 2013.

[9] Q. Wang, J. Zheng, H. Xu, B. Xu, and R. Chen, "Roadside magnetic sensor system for vehicle detection in urban environments," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 5, pp. 1365–1374, 2017.

[10] L. O. Chua and T. Roska, "The cnn paradigm," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 40, no. 3, pp. 147–156, 1993.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587, 2014.

[12] R. Girshick, "Fast r-cnn," in Proceedings of the IEEE international conference on computer vision, pp. 1440–1448, 2015.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, 2015.

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in European conference on computer vision, pp. 21–37, Springer, 2016.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016.

[16] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[17] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[18] B. Chougula, A. Tigadi, P. Manage, and S. Kulkarni, "Road segmentation for autonomous vehicle: A review," in 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 362–365, IEEE, 2020.

[19] J. Tang, S. Li, and P. Liu, "A review of lane detection methods based on deep learning," Pattern Recognition, vol. 111, p. 107623, 2021.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, pp. 234– 241, Springer, 2015.

[21] Z. Wang, W. Ren, and Q. Qiu, "Lanenet: Real-time lane detection networks for autonomous driving," arXiv preprint arXiv:1807.01726, 2018.

[22] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, 2018.

[23] Q. Zou, H. Jiang, Q. Dai, Y. Yue, L. Chen, and Q. Wang, "Robust lane detection from continuous driving scenes using deep neural networks," IEEE transactions on vehicular technology, vol. 69, no. 1, pp. 41–54, 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:48:32 UTC from IEEE Xplore. Restrictions apply.

[24] R. C. Staudemeyer and E. R. Morris, “Understanding lstm–a tutorial into long short-term memory recurrent neural networks,” arXiv preprint arXiv:1909.09586, 2019.

[25] K. Cho, B. Van Merrie¨nboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” arXiv preprint arXiv:1409.1259, 2014.

[26] K. He, G. Gkioxari, P. Dolla´r, and R. Girshick, “Mask R-CNN,” in Proceedings of the IEEE international conference on computer vision, pp. 2961–2969, 2017.

[27] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004.10934, 2020.

[28] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv preprint arXiv:2207.02696, 2022.

[29] D. Wu, M.-W. Liao, W.-T. Zhang, X.-G. Wang, X. Bai, W.-Q. Cheng, and W.-Y. Liu, “Yolop: You only look once for panoptic driving perception,” Machine Intelligence Research, vol. 19, no. 6, pp. 550–562, 2022.

[30] Y. Qian, J. M. Dolan, and M. Yang, “Dlt-net: Joint detection of drivable areas, lane lines, and traffic objects,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 11, pp. 4670–4679, 2019.

[31] D. Vu, B. Ngo, and H. Phan, “Hybridnets: End-to-end perception network,” arXiv preprint arXiv:2203.09035, 2022.

[32] C. Han, Q. Zhao, S. Zhang, Y. Chen, Z. Zhang, and J. Yuan, “Yolopv2: Better, faster, stronger for panoptic driving perception,” arXiv preprint arXiv:2208.11434, 2022.

[33] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in 2012 IEEE conference on computer vision and pattern recognition, pp. 3354–3361, IEEE, 2012.

[34] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Be nenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3213–3223, 2016.

[35] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 year, 1000 km: The oxford robotcar dataset,” The International Journal of Robotics Research, vol. 36, no. 1, pp. 3–15, 2017.

[36] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu,A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “nuscenes: A multimodal dataset for autonomous driving,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 11621– 11631, 2020.

[37] J. Xue, J. Fang, T. Li, B. Zhang, P. Zhang, Z. Ye, and J. Dou, “Blvd: Building a large-scale 5d semantics benchmark for autonomous driving,” in 2019 International Conference on Robotics and Automation (ICRA), pp. 6685–6691, IEEE, 2019.

[38] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, “Bdd100k: A diverse driving dataset for heterogeneous multitask learning,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2636–2645, 2020.

[39] R. Padilla, S. L. Netto, and E. A. Da Silva, “A survey on performance metrics for object-detection algorithms,” in 2020 international confer- ence on systems, signals and image processing (IWSSIP), pp. 237–242, IEEE, 2020.

[40] H. Wang, R. Fan, Y. Sun, and M. Liu, “Dynamic fusion module evolves drivable area and road anomaly detection: A benchmark and algorithms,” IEEE transactions on cybernetics, vol. 52, no. 10, pp. 10750–10760, 2021.

[41] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890, 2017.

[42] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, “Enet: A deep neural network architecture for real-time semantic segmentation,” arXiv preprint arXiv:1606.02147, 2016.

[43] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, “Erfnet: Effi- cient residual factorized convnet for real-time semantic segmentation,” IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 1, pp. 263–272, 2017.

[44] M. Teichmann, M. Weber, M. Zoellner, R. Cipolla, and R. Urtasun, “Multinet: Real-time joint semantic reasoning for autonomous driving,” in 2018 IEEE intelligent vehicles symposium (IV), pp. 1013–1020, IEEE, 2018.

[45] J. Lee, S. Walsh, A. Harakeh, and S. L. Waslander, “Leveraging pre-trained 3d object detection models for fast ground truth generation,” in 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 2504–2510, IEEE, 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:48:32 UTC from IEEE Xplore. Restrictions apply.

A single–Port to Dual-Port Reconfigurable 7T SRAM Bit Cell for High Speed, Power Saving and Low Voltage Application

Bhawna Rawat

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India , bhawnarawat12@gmail.com

Poornima Mittal

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India , poornimamittal@dtu.ac.in

The decreasing operational voltage and scaled technology node for memory designing has widened the gap between two crucial parameters for an SRAM – delay and power. As the demand for internet of things is increasing, the need for round the clock connectivity is increasing. This mandates designing a cell with a capability to switch between low power and high speed operation. Thus, this paper presents design of dual mode operational 7T cell that can switch between single port and dual port configuration. The proposed reconfigurable cell can operate as single port or dual port cell single ended 7T cell. The reconfigurability in the cell is realized using control signals. The noise stability of the bit cell is obtained to be 333, 333, and 470 mV for read, hold and write modes respectively. The robustness of the cell against temperature variation, process variation and voltage variation is also analyzed. The performance variation in each parameter will not have a dramatic impact as it is within manageable limit. Its write time is 0.14 ns, while 5 ps are required for a successful read operation. The dual port configuration of the cell supports pipelining and thus operates faster than its single port configuration.

CCS CONCEPTS: Hardware ~Integrated circuits ~Semiconductor memory ~Static memory

Additional Keywords and Phrases: Reconfigurable cell, low power, high speed, process tolerant, internet of things

1 INTRODUCTION

Static random access memory (SRAM) is a fundamental part of most multimedia systems [1-2], consequently, improving the performance of an SRAM based memory is crucial for performance enhancement of the setup. Additionally, with the surge in requirement of Internet of Things (IoT), the demand for high speed, low power, and portable devices has grown exponentially. According to the present statistics, over 70% multimedia devices utilize SRAM for cache memory design, which accounts for upto 30% of the total power consumption and 50% of die area for the system [3]. In keeping with the present multimedia device requirements, a memory designer has to ensure low power and high speed operation with high packaging density. The major concern for a memory designer is the dichotomic requirement of low power and high speed operation. The two demands are complementary and one may be achieved at the cost of the other. As a result, the gap between the two is pushing designers to search for alternatives to rectify the same. The problem is further intensified by the shrinking technology node to increase the integration density of the memory. The increasing demand for round the clock connectivity and burst towards IoT has generated the demand for versatile SRAM.

Additionally, an SRAM array is common for a system and is shared between circuits with different error tolerance, delay, and power requirements. For example, image processing application in a multimedia device has higher tolerance to error but requires faster operation [3]. Whereas, data centric applications need the memory performance to be error proof, while being slightly relaxed about the speed of the circuit [4]. But, this makes the system a bit power hungry. Thus, each system has different requirements – faster operation, lower power, high error tolerance, in keeping with the application that it is presently catering to.

But from a designer's prospective, error tolerance, low power consumption, faster operation, and high integration density are trade-offs [5]. Therefore, achieving them all at the same time is an extremely difficult task in a conservative SRAM cell design. Thus, a new concept of reconfigurable SRAM is described in this paper. The proposed cell is a versatile SRAM cell designed with the capability to function in two different single ended configurations – single port and dual port. The single port SRAM bit cell is power efficient but with higher delay and greater error propensity than the dual port cell [6]. While, the dual port cell is more effective for faster operation but is power hungry [7]. Also, the dual port cell enables pipelining, whereas this is not the case with single port cell. The paper is divided into the following five distinct sections, together with this introduction section. The concept of reconfigurable SRAM is explained in detail along with the previously reported scenarios for reconfigurability in section II. In section III, the structure of the proposed reconfigurable cell is explained in detail. The results of the reconfigurable cell are elaborated upon in section IV and its sub sections. The outcomes of the paper are summarized in section V.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
1084-4309/2023/1-ART1 $15.00
http://dx.doi.org/10.1145/3616872

Just Accepted

2 RECONFIGURABILITY IN SRAM

Reconfigurability in an architecture refers to its ability to rearrange the elements or settings of a system, device, or computer application. The concept of reconfigurability in devices has been around for a long time. Schuyler et al. [8] used reconfigurability to design memory cell that can function both as random access memory and read only memory. This helps to eliminate the need for designing an application specific test procedure. But, this patent only modulates the use of an already designed memory cell. Each cell in the memory block is not worked upon. Also, it is not able to improve the performance of the cell. Similarly, another patent that reports reconfigurability in memory was reported by Sutardja Sehat [9]. In it, the researcher has reported a block level memory structure, capable of working as per requirement of different application. The aforementioned patents describe a mechanism to use the existing memory cells to adapt as per the requirement of the application. But, neither of the two work on modifying the bit cell that stores the information. Whereas, the recent boom in demand for artificial intelligence and machine learning applications has generated the need for high performance memory circuits with minimum propensity for error. This demand is further aggravated by increased inclusive integration of different devices for round the clock connectivity via internet access for developing an IoT.

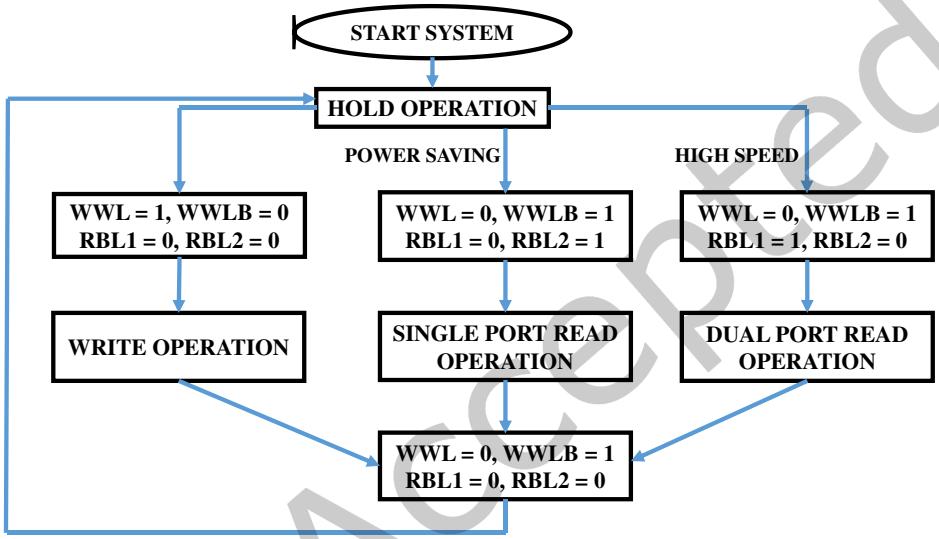


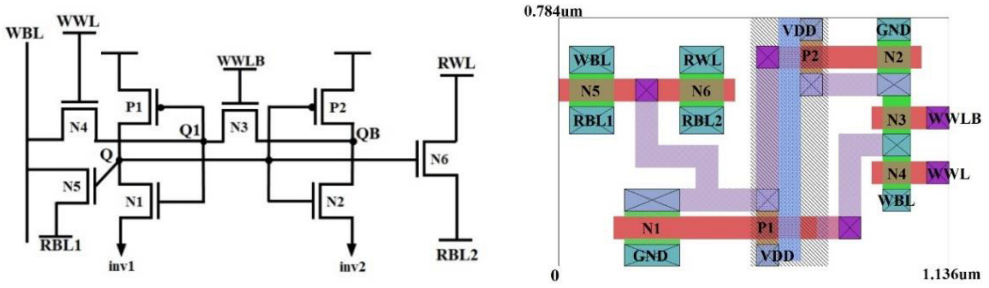
Fig. 1 Flow chart representation of the proposed reconfigurable SRAM

An IoT setup has fluctuation in its design requirement in keeping with the time of day. A device connected to the setup is expected to be operational at its full capacity for a fraction of the day; other times it may function in the power saving mode. The pre-existing bit cell designs are static and operate in only one mode. This limits the functionality for the bit cell. Whereas, to carter to the dynamic requirement for an IoT system, a bit cell that can operate in different modes is the requirement. In pre-exiting bit cells it is observed, that single ended single port bit cells are better for power saving application. Whereas, dual port cells perform better in terms of speed. As previously stated, an IoT setup is operational round the clock, but optimal performance is required in smaller bursts. This mandates designing a bit cell that can switch between two different modes of operation. Thus, in this paper, the concept of a reconfigurable SRAM bit cell is presented. Its working is explained with the help of a flow chart demonstrated in Fig. 1. The working of a reconfigurable cell is realized for a 7T SRAM cell. The structure of the reconfigurable cell is described in section III. This reconfigurable cell includes control circuit that enables the cell to either work in single port or dual port configuration. The bit cell is composed of eight transistors and is grouped into three sub-parts – single bit memory core, reading port and writing port. The single bit memory core of the reconfigurable memory is the part that stores the desired information. The read and the write port are the access circuitry that enable the device to read and write into the cell respectively.

3 DETAILED DESCRIPTION OF THE PROPOSED RECONFIGURABLE SRAM CELL

The proposed cell is composed of eight transistors of which two are PMOS (P1 and P2), while the remaining are NMOS transistors (N1-N6). The memory core for the reconfigurable cell alike a conventional SRAM cell is formed by a pair of mutually connected inverters via an NMOS transistor; N3. It is used to control the connection and disconnection of mutual feedback between inverters; it is controlled via signal write wordline bar (WWLB). While, the write wordline (WWL) signal controls the access transistor - N4 for enabling the write operation.

Just Accepted



(a) (b)

Fig 2. (a) Schematic design and (b) Layout for the proposed reconfigurable SRAM cell

The WWL and WWLB are complementary signals thereby, making N4 and N3 operate exclusive of each other. The read operation is controlled by read bitline - RBL1 or RBL2, signal depending on whether single or dual port configuration is to be used respectively. The N5 - N6 transistors are the read access transistors and operate exclusive of each other. The data is read from or written into the memory with the assistance of bitline WBL. The schematic design for the proposed reconfigurable SRAM cell is depicted in Fig. 2 (a). While, its layout along with dimensions is depicted in Fig. 2 (b). The layout for cell is designed on a grid of 16 nm and based on design rules reported by Jeppson *et al.* [10]. The area footprint for the proposed bit cell is found to be 0.891 μm^2.

Table 1. The control signal conditions during the different operation of the proposed cell

	WWL	**WWLB**	**RBL1**	**WBL**	**RBL2**	**RWL**
HOLD	'0'	'1'	'0'	'0'	'0'	'0'
WRITE	'1'	'0'	'0'	'0'	'0'	'0'
READ (SP)	'0'	'1'	'0'	'1'	'1'	'1'
READ (DP)	'0'	'1'	'1'	'1'	'0'	'1'

This proposed configuration for the bit cell can operate in two modes: (i) single-ended, single-port and (ii) single ended, dual port. The port configuration for a cell indicates if common bitline is used for both access operations or are the bitlines independent for the two operations. A single port bit cell relies on a single bitline for all its operations. Whereas, a dual port cell is dependent on two bitlines for its operations. In the proposed cell use of the control signal RBL1 for read operation ensures single port configuration and if RBL2 signal is used, then the circuit will operate in dual port configuration. For the ease of understanding, the combination of control signal and their corresponding operations are listed in Table 1.

3.1 Single Port Configuration

When a device is operating at a slow pace or is in low activity and power saving mode, then the circuit must operate in single port configuration. During this configuration, the bit cell is dependent on WBL for all its operations. The RBL2 signal is low, thereby disconnecting the second read port for the cell. The effective circuit for the proposed cell during the single port configuration is given in Fig. 3 (a). For this configuration, the cell operation is slow with no possibility for pipelining. But, use of a single ended single port configuration tends to reduce power consumption; mathematically calculated as follows:

$$P = \alpha C_L V_{DD}^2 f_{clk} \quad (1)$$

Here, α represents the activity factor, f_{clk} is the clock frequency and C_L is the load capacitance of the cell. For a single ended operation, the activity factor is reduced to 0.5 thereby, reducing the power appetite to half of its value.

Just Accepted

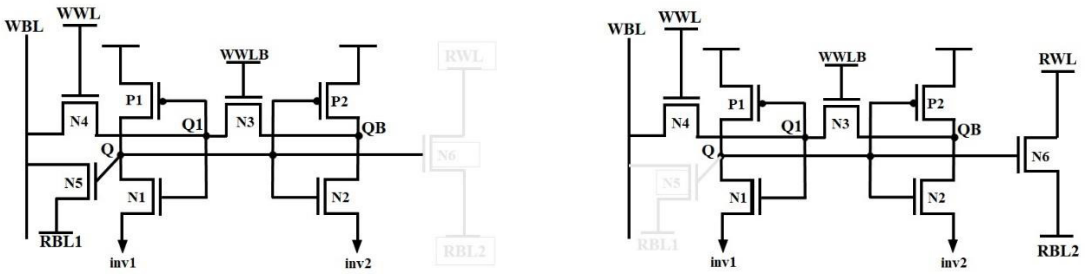


Fig 3. The effective circuit for the proposed cell when the circuit is operational in (a) single port, and (b) dual port configuration

3.2 Dual port configuration

When the device is operating in a high speed mode, i.e., the cell is required to operate with lower delay and higher speed. This scenario trades-off power with speed and is not taken as a constraint, then the cell must operate in dual port. The effective circuit for the proposed reconfigurable SRAM cell for dual port configuration is shown in Fig. 3 (b). The advantage of dual port is that it enables pipelining and ensures faster operation. This might increase the power load of the circuit but the operational speed of the cell is optimum.

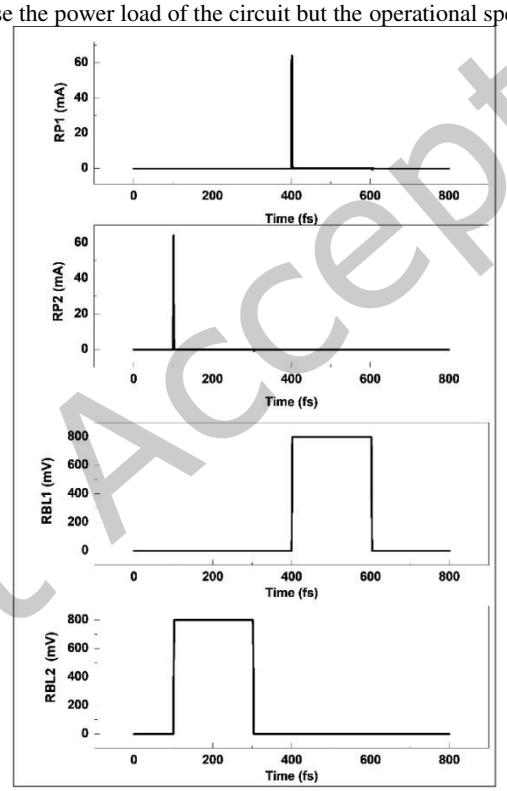


Fig. 4 Read current through the different read ports for the dual mode operational SRAM cell.

3.3 Functioning of proposed cell

The proposed cell is designed to operate in two different modes – single ended, single port and single ended dual port. It is essential to analyze the working of the cell in the two modes. Therefore, working of the proposed dual mode operational cell in two different read modes is demonstrated in Fig. 4. When the cell performs read operation, read discharge current flows through the selected read port – single or dual. Thus, the read operation for the bit cell is demonstrated in terms of read discharge current. The initial condition for the cell is maintained to be Q = '1', and QB = '0'. The write port for the cell is turned OFF and only the read operation is to be performed. It can be observed from Fig. 4, that when the RBL1 signal is high, the read current through read port1 (RP1) increases instantly to 64.65 mA.

This is the read discharge current that is responsible for the read operation. During this phase, the cell is being operated in single ended single port configuration, but as the read port2 (RP2) is also connected to node Q; a fickle leakage current can be observed at that port too. Whereas, when the cell is being operated in the dual port configuration, the read discharge current flows through RP2 only. This read discharge current is also in the same range but is slightly higher than the RP1 read current at 64.14 mA. Thus, based on the read discharge

Just Accepted

current recorded for the mode of operations for the cell, it is justified that it will be able to operate in dual mode configuration. Thus, as the proposed dual mode configuration cell may be deemed stable. Also, the cell can operate in both the single and dual port configurations independently.

Another important inference that can be drawn from Fig. 4 is that, small leakage current pulse is observed during the read operation in the read port that is not selected. When RBL1 is high, the read discharge current flows through RP1, and a leakage current of 189 µA is registered for RP2. Similarly, when the RBL2 signal is high, read current is observed through RB2, and leakage current of 193.2 µA is registered through RP1. The leakage current for a port is an indicator of its leakage power consumption. The leakage power consumption for port is determined as the product of supply voltage and leakage current. Thus, based on the leakage current values for RP1 and RP2, the leakage power consumption value for the cell is 151.2 and 154.56 µW, respectively.

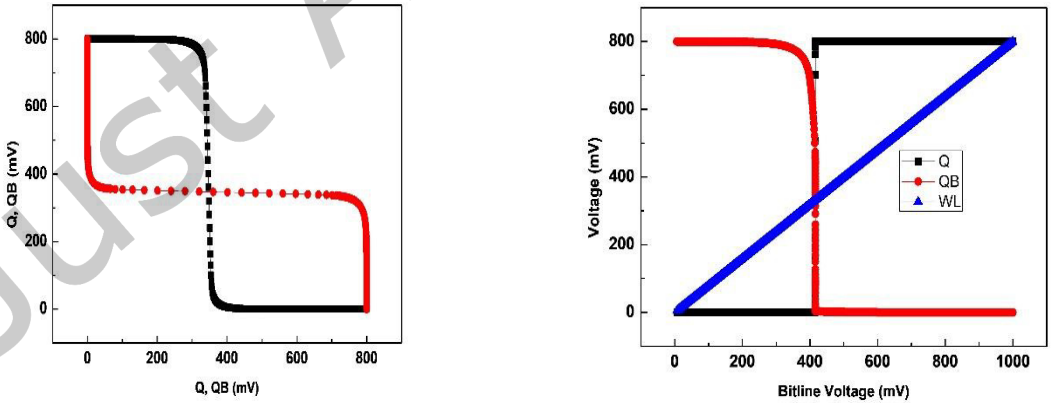
4 PERFORMANCE EVALUATION

The proposed cell in this paper is designed at 32 nm and is simulated at V_{DD} of 0.8 V. The different parameters used to evaluate the proposed reconfigurable cell are static noise margin, timing requirement for access operations (read and write), process, temperature, and voltage variation. The performance of the proposed reconfigurable cell for each of the aforementioned parameters is explained in the subsequent sub sections.

4.1 Static Noise Margin

The primary utility of a memory cell is to retain information as long as V_{DD} is applied to the cell. Therefore, the proposed reconfigurable cell is firstly analyzed for static noise margin (SNM) for its three operations - hold, read, and write. The performance is initially analyzed for ideal static condition. In an ideal condition, the SNM for an operation is defined as the utmost noise a cell can withstand before an error is registered in the stored data. The SNM values for the hold, read, and write are referred to as HSNM, RSNM, and WSNM respectively. The hold, read, and write SNM values obtained for the proposed reconfigurable SRAM cell are 333, 333, and 470 mV respectively. The butterfly curve (hold/read operation) and write margin curve are depicted in Fig. 5 (a) and (b) respectively.

Overlapping butterfly curves are obtained for read and hold operation; this is an implication of the read SNM free port for the proposed cell. The read ports (for both single port and dual port configuration) for the cell are designed to exclude the data node from the read discharge path. This configuration helps protect the integrity of the data stored in cell during the read operation. Thus during the read operation, the data stored at node Q is not disturbed. As the data remains undisturbed, the RSNM value for the cell is equal to its HSNM value of the cell. Such a read port configuration is referred to as a read SNM free port [7].



(a) (b)

Fig. 5 (a) Butterfly curve for hold (read) operation, and (b) Write margin curve for the proposed reconfigurable cell.

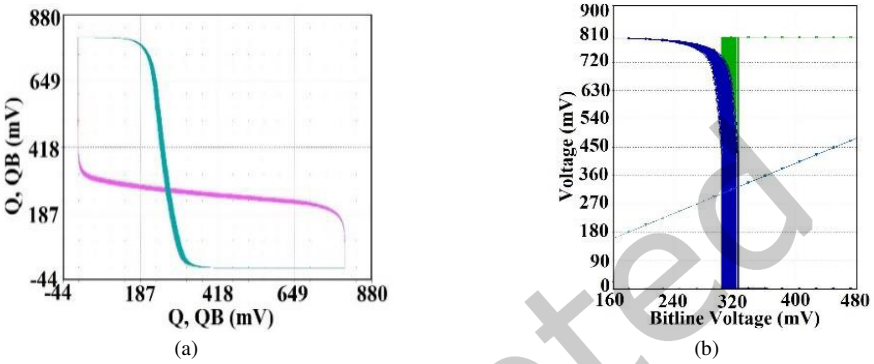
The HSNM (RSNM) value for the proposed cell is 333 mV, while the WSNM value for the cell is 470 mV. The SNM values obtained for the cell are fairly balanced and lie around the ideal value of $V_{DD}/2$ [11]. But, when the memory is fabricated, imperfections in the cell may arise due to different process variables. These imperfections become more prevalent at lower technology node. The different factors that impact a cell's performance and threaten its stability are local and global variations caused due to fabrication imperfections.

4.2 Local Variation

The simulation results for the proposed cell demonstrated in the previous sub section are all performance for ideal conditions. But, when a circuit is fabricated, there are possibilities of minor fabrication imperfections resulting in local mismatch in the transistors adjacent to each

Just Accepted

other. This results in deviation from the performance reported for the cell. Thus, it is essential to evaluate the impact of local variation on the performance of the proposed cell. Factors that limit the performance of the bit cell are; hot carrier injection, gate oxide breakdown, backend dielectric breakdown, electro migration, and stress-induced voiding. But these factors have a random behavior and do not have a defined model of analysis. All variations can be translated to an effective variation in the threshold voltage (V_{TH}) [12, 13]. Consequently, to analyze the impact of variation on device, the Gaussian distribution for V_{TH} is an effective approach [14]. Thus, V_{TH} is used as the main parameter to statistically analyze the same [15-16]. Monte Carlo simulation is employed to determine degree of variation in performance of the proposed cell. The V_{TH} for the cell is varied around the average value with 6σ standard deviation to determine the worst case performance of the proposed reconfigurable bit cell. The results for the Monte Carlo simulation corresponding to the hold (read) and write operation for the proposed cell are depicted in Fig. 6 (a) and (b), respectively.



(a) (b)

Fig. 6 Results of local variation for (a) HSNM (RSNM), and (b) WSNM values for the proposed reconfigurable cell.

The variation in the butterfly curve of HSNM (RSNM) for the reconfigurable cell is demonstrated in Fig. 6 (a), while the variation in the WSNM is represented in Fig. 6 (b). The 6σ worst case HSNM, RSNM, and WSNM values achieved for the reconfigurable cell are 273, 273, and 540 mV respectively. Thus, the variation in HSNM, RSNM and WSNM values for the proposed cell are 60, 60, and 70 mV respectively. The maximum impact of local variations and global variations (explained in the subsequent section) on the performance of the proposed cell is depicted in Fig. 8. From the local variation analysis, this may be concluded that though there is variation in performance of the bit cell, it is within reasonable bounds. Therefore, it is apt to suggest that the cell has low vulnerability to local variation.

4.3 Global variation (Process corner analysis)

When any inorganic circuit is fabricated, there are various factors such as – oxide thickness, strain variation, line edge roughness, and discrete random dopant amongst others that may cause process variations in the fabricated circuits [17]. The impact of these variations is more prevalent in circuits designed for scaled technology nodes. Therefore, the proposed reconfigurable bit cell is prone to error caused by global variations. Thus, it is essential to check the range of variation in cell performance due to global variation. The impact of global variation on a cell is determined using the process corner analysis. The proposed cell is therefore analyzed at the four process – Fast-Fast (FF), Fast-Slow (FS), Slow-Fast (SF), and Slow-Slow (SS) along with the typical corner (TT). The SNM values obtained for the proposed cell at all the process corners are graphically compared in Fig. 7 and the corresponding values are tabulated in Table 2.

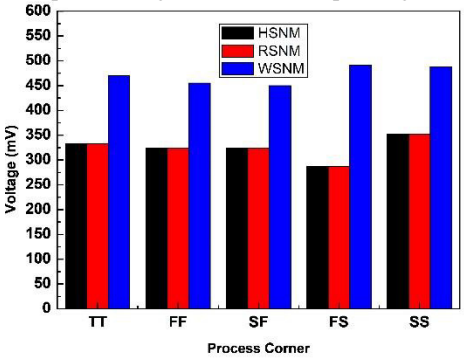


Fig. 7 Graphical comparison of the SNM values obtained at the different process corner for the proposed reconfigurable cell.

Just Accepted

Table 2 SNM values for the proposed reconfigurable cell at all process corners

	TT	FF	FS	SF	SS	Maximum Variation
HSNM	333	340	352	287	324	65
RSNM	333	340	352	287	324	65
WSNM	470	455	450	491	488	41

The best performance for the cell is recorded at the FF corner. The HSNM, RSNM, and WSNM values obtained at the FF corner are 340, 340, and 455 mV, respectively. This is an implication of improved performance for both NMOS and PMOS at the FF corner. The deviation in the HSNM, RSNM, and WSNM values obtained at the FF corner in comparison to values obtained at the typical corner is 7, 7, and 15 mV, respectively. At the typical corner, the performance of the PMOS is weaker in comparison to an equally sized NMOS. Thus, when the cell is functioning at the FS corner, the gaps between the performance of PMOS and NMOS transistors is bridged, thus the best performance for the cell is recorded with the HSNM, RSNM, and WSNM being 352, 352, and 450 mV, respectively. These values are deviated from the typical corner values by 19, 19, and 20 mV, respectively.

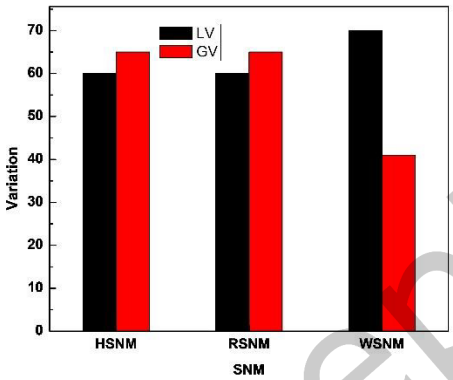


Fig. 8 The variation in performance of HSNM, RSNM, and WSNM due to local and global variation

Contrary to the FS corner, the gap between the performance for PMOS and NMOS is further widened at the SF corner. Consequently, the performance of the proposed cell is also most inferior at this corner. The SNM values recorded for the hold, read, and write operation at this corner are 287, 287, and 491 mV, respectively. These values are 46, 46, and 21 mV deviated from the typical corner values. The performance for the proposed cell at the SS corner is moderate, with the HSNM, RSNM, and WSNM values being 324, 324, and 488 mV respectively. The essence of the global variation analysis is to determine the variation in range of SNM values. The range for SNM variation is the difference between the highest and lowest value obtained. Thus, the maximum deviation in the HSNM, RSNM, and WSNM values due to global variation is 65, 65, and 41 mV respectively. Thus, the proposed reconfigurable cell may be deemed resistant to global variation as it registers a fairly small change in its performance due to global fabrication imperfections.

4.4 Temperature Variation

Local and global variations in a circuit are an implication of fabrication imperfections. But, external factors such as temperature variation also have a highly likely possibility of impacting the performance of a cell. Therefore, to evaluate robustness of the proposed cell against temperature variation, its static performance is analyzed for -10 ºC to 110 ºC. The variation in HSNM/RSNM and WSNM curves for the proposed reconfigurable cell against temperature variation is depicted in Fig. 9 (a) and (b) respectively.

Just Accepted

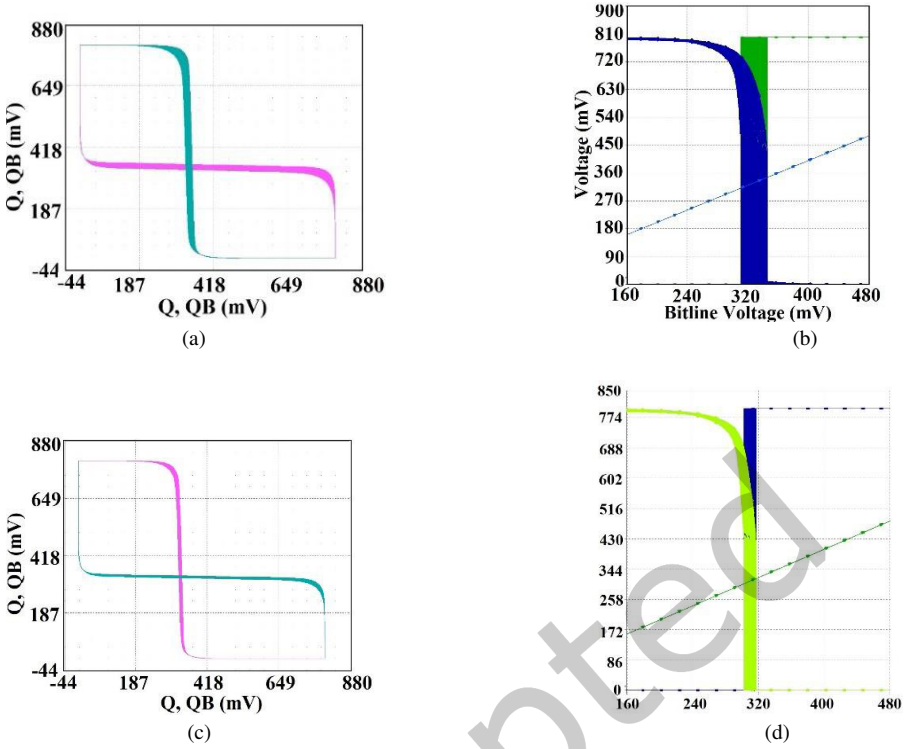


Fig. 9 Variation in SNM for proposed cell during (a) hold (read), and (b) write operation for temperature variation -10 ºC to 110 ºC, and variation in SNM during (c) hold (read), and (d) write operation for temperature variation 0 ºC to 70 ºC

The proposed reconfigurable cell records a small variation in its performance when temperature is varied. The lowest HSNM (RSNM) value obtained for the proposed cell due to temperature variation is 308 mV. This is a 25 mV deviation from the nominal SNM value at room temperature. The WSNM value for the cell rises to 520 mV. This is a 50 mV rise from the WSNM value at 27ºC. But, this temperature range taken into consideration is extremely large between -10ºC to 110ºC. Whereas, range for temperature variation in most commercial electronic devices is 0ºC to 70ºC. Thus, the proposed cell is evaluated for this range. The HSNM (RSNM) and WM curve for temperature variation in commercial range are depicted in Fig. 9 (c), and (d) respectively. It can be inferred from Fig. 9 (c) and (d), that the variation in performance of the proposed cell is considerably reduced for the commercial temperature range. The maximum variation in HSNM (RSNM) and WM for the commercial temperature range is 28, and 24.5 mV, respectively. Thus, it may be concluded that the proposed reconfigurable cell demonstrates reasonable performance against temperature variation.

4.5 Voltage Variation

Digital circuits operate for bit values ‘0’ and ‘1’, but bit levels 0 and 1 in reality are not discrete level; instead they are defined as voltage range. Consequently, it is essential to analyze the cell performance for voltage variation. In this work, the proposed reconfigurable cell is analyzed for ±10% of the operational V_{DD}. The influence of voltage variation on cell stability is depicted in terms of deviation in its HSNM (RSNM) and WSNM values. The simulation results for voltage variation analysis are presented in Fig. 10.

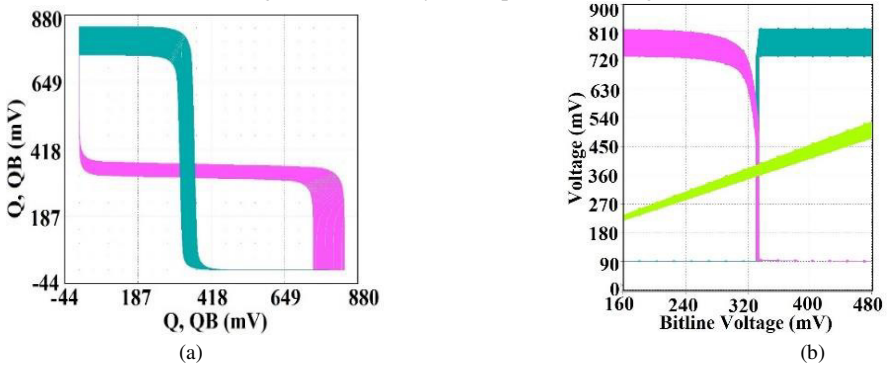


Fig. 10 Variation in SNM values for (a) hold (read) and (b) write operation for the proposed cell due to variation in voltage from 0.75 V to 0.85 V.

Just Accepted

Table 3 SNM values for the proposed reconfigurable cell for different V_{DD} values

V_{DD} (V)	HSNM (mV)	RSNM (mV)	WM (mV)
0.8	333	333	470
0.7	286.6	286.5	414
0.6	235.3	235.3	362
0.5	188	188	309
0.4	144	144	256
0.3	96	96	204

It may be observed from Fig. 10 that voltage variation has larger impact on the cell performance in comparison to process and temperature variations. But, the variation in cell performance is untroublesome and therefore, have no dire consequence on the performance of the cache memory. Additional to voltage variation analysis, as there is a tremendous growth in devices operational at low voltage, it is essential to analysis the performance of the proposed dual port cell for the same. Therefore, the SNM values for the proposed cell are analyzed and recoded for V_{DD} lower from 0.8 V to 0.3 V in step size of 0.1 V. The results obtained for the analysis are tabulated in Table 3. It may be inferred from the values tabulated in Table 3, that as the V_{DD} is lowered the stability of the cell also reduces linearly. The HSNM (RSNM) value decreases approximately 47 mV per 100 mV decline in V_{DD}. Whereas, for WM, for each 100 mV decline in V_{DD} the WM value reduces approximately 52 mV. Thus, it may be inferred that the cell will be able to retain its performance for low values of V_{DD} as well.

4.6 Write and Read Time Analysis

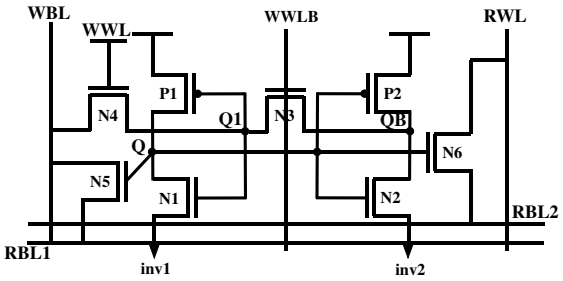
The analysis explained thus far are all static in nature and do not take into account the time needed for successful completion of an operation. But, a SRAM cell is a peg in a complex system on chip setup. Thus, identifying the minimal time required to perform an operation is essential for a cell's performance evaluation. The write time for an SRAM bit cell is defined as the time interval between the instant when the write word line signal is exerted and the data stored in the cell is reversed [18]. The proposed reconfigurable cell requires a minimum pulse of 0.14 ns to successfully perform the write operation. Both the node values attain their respective high/low level before the word line signal is exerted low.

Another operation that is highly time dependent is the read operation. Both the read ports for the reconfigurable cell are single ended and identical in configuration. Due to its read SNM free configuration and single ended nature, the read discharge current is observed only for Q = '1', RWL = '1' and RBL1/RBL2 = '0'. For node Q = '0' the read port NMOS transistor N5/N6 (single/dual port configuration) remains OFF and no read current event is registered. The read time for the cell is estimated as the time required to lower the voltage of the bitline to the switching voltage of an inverter [19]. The read time for the cell was identified to be 5 ps and in ideal condition is identical. But, single port cells do not support pipelining, whereas, dual port cells enable it. Thus, when the memory is required to operate in high speed mode, the cell should be steered into dual port configuration. Whereas, when timing is not a constraint, the cell may function in its single port mode.

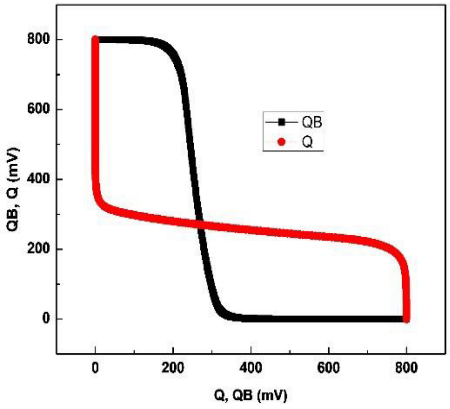
4.7 Half Select Disturbance

The bit cell thus far has been analyzed in isolation. But, a bit cell is part of a large array, where cells in the same row and column share common signals. Thus, when signals are exerted to select a particular cell, other cells in the same row as well as column get half selected. These half selected cells may erroneously register an unintentional read or write operation. This phenomena is referred to as the half select disturbance (HSD) [20, 21]. It is a common flaw in memories with share wordline architecture [22]. The conventional and successful mechanism to avoid HSD is to route control signals for an operation that are exclusively row and column based [23].

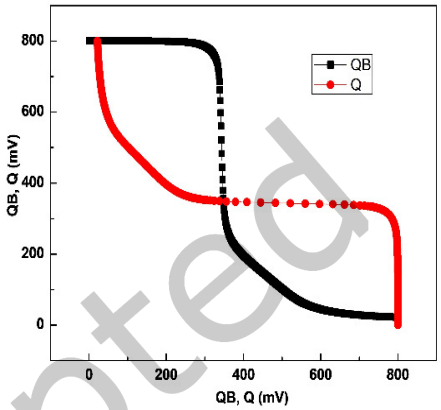
Just Accepted



(a)



(b)



(c)

Fig. 11 (a) Schematic diagram for the proposed cell representing routing of control signals to avoid HSD, butterfly curve for half selected cell when (b) WWL is high and, (c) WWLB is low.

To ensure that the proposed cell is HSD free, the signals for read and write operation are exclusively routed to be row and column based. The same is demonstrated in Fig. 11 (a). The write control signals WWL and WWLB are routed as row and column based signal respectively. Similarly, for the single port read operation WBL and RBL1 are routed mutually exclusive row and column based. The same is the case for RWL and RBL for dual port read operation. The read operation for the cell is SNM free, thus its stability is secure during the read operation. Thus during the half select condition, even if the cell is partially select for read, the cells stability would not be disturbed. But, during the write operation it is essential to check for cell stability. Thus, if a cell in the same row is selected for write operation the WWL signal will be high. Whereas, if the selected cell for write operation is in same column, the WWLB signal will be low. Corresponding to both the cases the stability curve for the proposed cell is depicted in Fig. 11 (b), and (c) respectively. The HSNM for the cell reduces for WWL = '1'to 250 mV, and for the second case of WWLB = '0', the HSNM reduces to 277 mV. Thus, it may be inferred that the stability of the proposed cell may suffer slightly due to HSD, but its performance is reasonable.

5 CONCLUSION

The growing dependence of humans on intelligent systems for their day to day activities has fueled the need for low power, high speed, and energy efficient devices. The key element of these intelligent devices is a high performance cache memory. But, the decreasing operational supply voltage and scaled technology node for memory designing has widened the gap between two crucial trade-off for an SRAM cell – delay and power. The need for round the clock connectivity is increasing with the growing popularity of internet of things. This mandates designing a cell with a capability to switch between low power and high speed operation. Thus, a reconfigurable 7T SRAM cell is proposed in this paper. The proposed reconfigurable cell can operate as a single port or dual port single ended 7T memory cell. The switching of the mode of operation of the bit cell is controlled using the control signals for the cell. The noise stability of the bit cell is measured to be 333, 333, and 470 mV for read, hold and write operation, respectively. The robustness of the cell against temperature variation, process variation and voltage variation is also analyzed. The variation recorded in each performance parameter is within acceptable limit. Therefore, is highly likely to not have any radical impact on the overall performance of the cell. The write time is recorded to be 0.14 ns, while 5 ps is required to complete a successful read operation. The dual port configuration of the cell supports pipelining and thus, operates faster than its single port configuration.

Just Accepted

REFERENCE

[1] S. Yang, W. Wolf, N. Vijaykrishnan, "Power and performance analysis of motion estimation based on hardware and software realizations," *IEEE Transactions on. Computers*, vol. 54, no. 6, pp. 714–716, 2005.

[2] G. Chen and M. Kandemir, "Optimizing address code generation for array-intensive DSP applications," in *Proc. Int. Symp. Code Generation Optimization*, pp. 141–152, 2005.

[3] N. Surana and J. Mekie, "Energy Efficient Single-Ended 6-T SRAM for Multimedia Applications," *IEEE Transactions on Circuit Systems II: Express Briefs*, vol. 66, no. 6, pp. 1023-1027, 2018.

[4] B. Rawat, and P. Mittal, "A 32 nm single ended single port 7T SRAM for low power utilization," *Semiconductor Science and Technology*, vol. 36, no. 9, pp. 095006-095022, 2021.

[5] D. Dutt, P. Mittal, B. Rawat, and B. Kumar, "Design and Performance Analysis of High Low Power Voltage Mode Sense Amplifier for Static RAM", *Advances in Electrical and Electronic Engineering*, vol. 20, no. 3, pp. 285-293, 2022.

[6] B. Rawat and P. Mittal, "Single bit line accessed high performance ultra-low voltage operating 7T SRAM bit cell with improved read stability," *International Journal of Circuit Theory and Application*, vol. 49 no. 5, pp. 1435-1449, 2021.

[7] J. Singh, S.P. Mohanty and D.K. Pradhan, "Robust SRAM designs and analysis," New York: Springer, 2013.

[8] S.E. Shimanek, E.E. Edwards, T.J. Davies, "Reconfigurable SRAM-ROM Cell", US 7,023,744 B1, 4 Apr 2006

[9] S. Sehat, "System on chip with reconfigurable SRAM", TW200901225A, 1 Jan 2009

[10] K.O. Jeppson, S. Christensson, N. Hedenstierna, "Formal Definitions of Edge-Based Geometric Design Rules," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 12, no. 1, pp. 59-69, 1993.

[11] B. Rawat, P. Mittal, "A comprehensive analysis of different 7T SRAM topologies to design a 1R1W bit interleaving enabled and half select free cell for 32 nm technology node", *Proceedings of the Royal Society A: Mathematical, Physical, and Engineering Sciences*. vol. 478, pp. 20210745-20210771, 2022.

[12] A. Agarwal, B.C. Paul, S. Mukhopadhyay, and K. Roy, "Process variation in embedded memories: failure analysis and variation aware architecture," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1804, 2005.

[13] Y. Wang, S. Cotofana, and L. Fang, "A unified aging model of NBTI nd HCI degradation towards lifetime reliability management for nanoscale MOSFET circuits," *2011 IEEE/ACM International Symposium on Nanoscale Architectures*, 175, 2011.

[14] S. Pal and A. Islam, "9T SRAM Cell for Reliable Ultralow Power Applications and Solving Multi-Bit Soft Error Issue," *IEEE Transactions on Device and Material Reliability*, vol. 16, no. 2, pp. 172, 2016.

[15] S. Ahmad, M. K. Gupta, N. Alam, and M. Hasan, "Single-ended Schmitt Trigger based robust low power SRAM cell," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20 no. 8, pp. 2634-2642, Aug. 2016.

[16] M. Moghaddam, M.H. Moaiyeri, M. Eshghi, "Ultra low-power 7T SRAM cell design based on CMOS," *2015 23rd Iranian Conference on Electrical Engineering*, (2015).

[17] A. Agarwal, B.C. Paul, S. Mukhopadhyay, and K. Roy, "Process variation in embedded memories: failure analysis and variation aware architecture," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1804, 2005.

[18] S. Ahmad, M. K. Gupta, N. Alam, and M. Hasan, "Single-ended Schmitt Trigger based robust low power SRAM cell," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20 no. 8, pp. 2634-2642, Aug. 2016.

[19] S. Lin, Y.B. Kim, F. Lombardi, "Design and analysis of a 32 nm PVT tolerant CMOS SRAM cell for low leakage and high stability," *Integration*, vol. 43, no. 2, pp. 176–187, 2010.

[20] R. C. Baumann, "Soft errors in commercial integrated circuits," *International Journal on High Speed Electronics and Systems*, vol. 14, pp. 299–309, 2004.

[21] Y.W. Chiu, Y.-H., Hu, M.-H. Tu, J.-K. Zhao, Y.-H., Chu, S.-J. Jou and C.-T. Chuang, "40nm bit-interleaving 12T subthreshold SRAM with data-aware write-assist," *IEEE Transactions on Circuits Systems I: Regular Papers*, vol. 61, pp. 2578–2585, 2014.

[22] A. Sachdeva, and V.K. Tomar "Design of Low Power Half Select Free 10T Static Random Access Memory cell," *Journal of Circutis, Systems and Computers*, vol. 30, no. 4, 2021

[23] D. Nayak, D. P. Acharya, P.K. Rout, and U. Nanda, "A high stable 8T-SRAM with bit interleaving capability for minimization of soft error rate," *Microelectronics Journal*, vol. 73, pp. 43, 2018.

Just Accepted

TechRxiv
Powered by IEEE

A Statistical Analysis Model of Wind Power Generation Forecasting for the Western Region of India

This paper was downloaded from TechRxiv (https://www.techrxiv.org).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

30-05-2023 / 14-07-2023

CITATION

Mahata, Sulagna; Harsh, Piyush; Shekher, Vineet; Rai, Pankaj (2023). A Statistical Analysis Model of Wind Power Generation Forecasting for the Western Region of India. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.23260151.v2

DOI

10.36227/techrxiv.23260151.v2

A Statistical Analysis Model of Wind Power Generation Forecasting for the Western Region of India

Sulagna M., Piyush H., Vineet S., and Pankaj R.

Abstract*— Wind energy has witnessed an upswing, with the improvement in current technology and cost-effective electricity production but due to uncertain wind behaviour and weather trends, it is essential to forecast wind energy. The paper presents the Autoregressive Integrated Moving Average model, a statistical analysis model, to forecast future power generation values, for the Kutch region of Gujarat, India. The historical wind power generation data and weather parameters have been taken into consideration for the predictive analysis of future trends in power generation. Wind power generation is dependent on various weather factors like wind speed, wind direction, temperature, humidity, air density, etc. The historical data obtained from Central Electricity Authority (CEA), India, and weather data collected from regional weather stations have been made into use for the forecast. A summary of the results is shown using the performance metrics for model evaluation, indicating that the model can forecast wind power generation with higher accuracy.

Keywords*— ARIMA, Power System, Renewable Energy, Wind Power forecasting

I. Introduction

Depleting natural resources and increasing global pollution have wheeled the development of the Renewable Energy industry. Human evolution has been fuelled by fossil fuels which are exhaustible and finite in number. The use of these non-renewable fuels has now become a major threat to the existence of mankind. The introduction of renewable energy sources instead of coal and oil-fuelled power plants has provided an alternative to conventional energy sources. Energy is vital for the economic and infrastructural development of the country. With the increasing demand for electricity in a developing country like India, energy sources like Solar Power, Wind Energy, and Geothermal Energy sources are the highlights for the days to come.

The Indian economy and growth have been so far anchored by fossil fuels. Recently, the wind power sector has witnessed a tremendous growth rate providing a key opportunity for the expansion and its integration into the power grid. With reference to the report by the Ministry of New and Renewable Energy, Govt. of India, India produces 42.633 GW of wind power making it the four largest installed wind power capacity in the world [1]. Accounting for about 10% of India's total installed energy generation capacity, Wind energy is India's oldest developed renewable energy technology [2]. More than 800 wind monitoring stations have been installed nationwide by the Indian government through the National Institute of Wind Energy (NIWE), which has also produced wind potential maps at elevations of 50, 80, 100, and 120 metres above ground level. According to the most current study, the country has a total wind power potential of 302 GW at 100 metres and 695.50 GW at 120 metres above ground level [1].

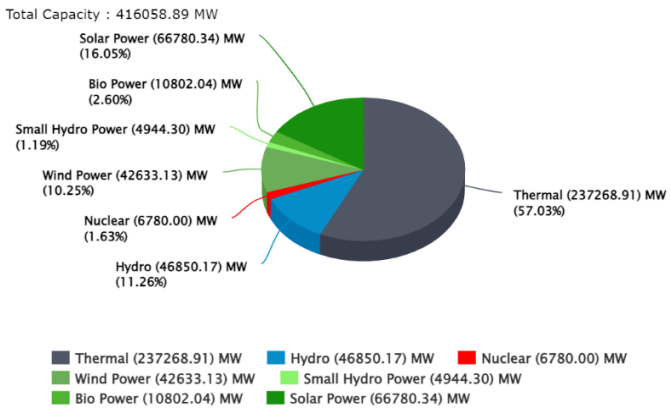


Fig. 1. Generation Category-wise Installed Capacity (as of 18/04/2023) [3]

Studies rank the state of Tamil Nadu with the highest installed capacity, which is well ahead of Gujarat which ranks second in terms of installed capacity. But in terms of potential energy generation, Gujarat ranks first with 142.56 GW (at 120 m AGL) [4].

Agricultural and industrial activities demand an extensive energy supply which the country can be with fulfil through a properly balanced energy supply through renewable energy and conventional energy sources. Thus, an precise prediction is needed to strategize, plan, and to schedule the power generation for smooth integration with the current power grid. A forecast of anticipated load requirements is essential for power system expansion, for determining generation capacity, transmission, etc. Smart management and the inclusion of renewable energy are important in today's scenario to meet the increasing demands. Wind power forecasting has been a focal point for researchers, hailing to its significance in energy balance in the grid and for proper scheduling of power generation.

Wind power forecasting methods broadly are of two types which include the physical method of analysing the physical quantity to obtain wind speed and then calculating the power thereafter, and the statistical method for establishing wind power prediction models by collecting historical data and training the model to yield future results [5].

Load forecasting is categorised into the following forms based on the forecasting horizon: Short-term (STLF), medium-

term (MTLF), and long-term load forecasting (LTLF). While the STLF technique typically spans time frame between an hour and one-week, medium-term load forecasts range from a few weeks to months and are considered a vital step in the electric power system. Long-term projections typically range from one year to 10 years in length [6] [7]. While LTLF is essential for the planning of power systems, MTLF is crucial for controlling fuel supply and maintenance [8]. STLF is important for daily operations of power systems such as energy trading in unregulated markets and unit dispatching or energy management [9].

A. Literature Review

In literature, time series forecasting methods are based on the analysis of historical data to predict future data points [10]. Load forecasting can be defined as the assessment of the amount of electricity needed by the customer and the utility systems. Various methods have been developed and deployed for forecasting. Average Meteorological Forecast (AMF) models are numerical weather prediction models, used to make predictions of wind speed are produced and then utilised to calculate wind power. Spatial Correlation Models make use of spatial relationships of wind speed of different sites [11]. Statistical models include models such as Auto-Regressive (AR), Auto-Regressive Moving Average (ARMA), and Auto-Regressive Integrated Moving Average (ARIMA), to find the correlation between on-line measured power data with off-line [10].

Adrian-Nicolae Buturache et al. presented a well-tuned ANN model to accurately forecast wind energy predictions but required time and computational power to find the optimum hyperparameter combination [12] Regression Tree models such as SVR provided transparency in comparison to ANN which required fine-tuning and optimization for the best results. But these tree-based models performed best when considering a single metric [12]. Wind energy is a very reliable energy source since weather factors crucially determine energy production. So, practically single metric models are not feasibly correct. Jung et al. reviewed the physical, statistical, spatial correlation, and regional forecasting models for wind speed and wind power forecasting [13].

On the basis of historical observations, Ernesta Grigonyte et al. applied ARIMA to forecast wind speed for a short-term period, considering the Baltic region. The authors tried to establish an optimal model, but a season-specific structure could not be attained [14]. Eddie Yatiyana et al. focused on forecasting wind speed and wind direction using a week of data from Western Australia based on the ARIMA model but could not perform the prediction using a single model to integrate wind speed and wind direction whereby the scenario becomes hypothetical [15]. Ilham Tyass et al. made predictions based on he Seasonal ARIMA model for the Morocco region. They obtained a forecast error of about 16% on an hour ahead forecast [16]. Similarly, Bri-Mathias Hodge et al. [17], Jing Shi et al. [18] and Pei Du et al. [19] addressed short-term wind power forecasting using statistical models, obtaining a similar trend in results with about 10% error percentage.

B. Location Analysis

Analysing the wind potential regions for energy generation in India and locating potent areas of Gujarat, this paper is focused on forecasting wind power generation in the Kutch region of Gujarat. Features like wind speed, temperature, humidity, and air density in the region favour the establishment of wind farms in this region. According to the National Renewable Energy Laboratory (NREL), the annual average wind speed concentrated in the region of the Gulf of Kutch and on the southern part of the peninsula is more than 8 m/s [20]. Also, due to the proximity of the Arabian Sea, the region has a higher air density which in turn helps in the rotation of the blades. Local air density changes with altitude, pressure, and temperature, thus directly affecting power output. Denser air exerts more force on the rotors, increasing power output.

C. Organisation of the paper

The authors are focused on long-term forecasting since the output would help strategize and plan for future expansion projects. Section I introduces the topic of interest and existing publications in the field, whereas the remainder of the paper is structured as follows: – Section II establishes the methodology i.e., the ARIMA model; in Section III, the experimentation and evaluation are discussed; and Section IV concludes the work and discusses the result.

II. Problem Formulation

The paper implements an optimized statistical model, ARIMA in conjunction with the Nelder-Mead optimization method after a comparative analysis with BFGS and L-BFGS optimisation methods.

A. Conceptualisation of Dependencies

The generation of wind power is significantly impacted by wind speed, wind direction and air density. The wind power can be determined using (1).

$$P = \frac{1}{2}\rho A V^3 \qquad (1)$$

Where,
P is the power in Watts,
ρ is the density of air in kg/m^3,
A is the cross-sectional area of the wind in m^2,
V is the velocity of the wind in m/s.

However, when calculating the wind power output, the coefficient of rotor performance is multiplied by the wind power obtained.

$$P = \frac{1}{2}\rho A V^3 C_p \qquad (2)$$

Where, C_p is the rotor coefficient of performance.

The coefficient of performance depends on the particular turbine and varies non-linearly with wind speed. The power a wind turbine can produce depends on the density of the air, the area swept by the turbine blades and the wind velocity. Among these, wind speed is a prominent factor, affecting power generation.

Air density also affects wind power generation, which is a function of pressure and temperature. In a real-world scenario, the air always consists of moisture.

$$\rho_{humid} = \frac{P_d}{R_d.T} + \frac{P_v}{R_v.T} \qquad (3)$$

Where,

ρ_{humid} is the density of humid air,
P_d is the partial pressure of dry air,
P_v is the partial pressure of water vapour,
T is the temperature (in K),
R_d is the specific gas constant for dry air,
R_v is the specific gas constant for water vapour.

B. Time Series Forecasting with ARIMA model

Time series forecasting models differ from traditional models since historical data is used to gauge future predictions. The proposed model for the forecasting of wind power is ARIMA which is a statistical model for non-stationary time series, widely used for short-term predictions. The notion of ARIMA was first developed by George Box and Gwilym Jenkins in their work "Time Series Analysis: Forecasting and Control" released in 1970 [21]. Box and Jenkins developed the ARIMA model as an extension of the autoregressive moving average (ARMA) model [14].

ARIMA models are based on the autoregressive (AR), moving average (MA), and differencing (I) concepts. The AR component simulates the linear dependence on its current and previous values. The error term is modelled by the MA component as a linear combination of previous error terms while the I component deals with the non-stationary nature by differencing it until it becomes stationary. p is the order of the autoregressive model (number of time lags), d is the degree of differencing (the number of times the data have had previous values subtracted), and the order of the moving-average part is denoted by q.

Mathematically, Autoregressive (AR) models are formulated as:

$$\gamma_t' = C + \phi_1\gamma_{t-1}' + \phi_2\gamma_{t-2}' \ldots + \phi_p\,\gamma_{t-p}' + e_t \qquad (4)$$

And Moving Average (MA) components are formulated as follows:

$$\gamma_t' = C + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} \ldots + \psi_q\,\varepsilon_{t-q} + e_t \qquad (5)$$

Where,

γ_t' where is the differenced series (it may have been differenced more than once),
'p' is the auto-regressive component,
'q' is the moving-average component,
e_t is white noise,
γ_{t-1}', γ_{t-2}', …, γ_{t-p}' denote the value of the variable at previous time periods, and
ε_{t-1}, ε_{t-2}, …, ε_{t-q} denote the error terms at previous time periods.

Combining both AR and MA parts, a typical ARIMA model denoted as ARIMA (p, d, q), is mathematically written as follows:

$$\gamma_t' = C + \phi_1\gamma_{t-1}' + \cdots + \phi_p\,\gamma_{t-p}' + \psi_1\varepsilon_{t-1} + \cdots + \psi_q\,\varepsilon_{t-q} + e_t \qquad (6)$$

In this work, the standard ARIMA model approach has been used for forecasting wind power generation. Based on the data records gathered, an optimised model was established using the Nelder-Mead optimisation method in conjunction with the estimation of parameters in the ARIMA model and compared with other ARIMA models with optimization methods, BFGS and LBFGS.

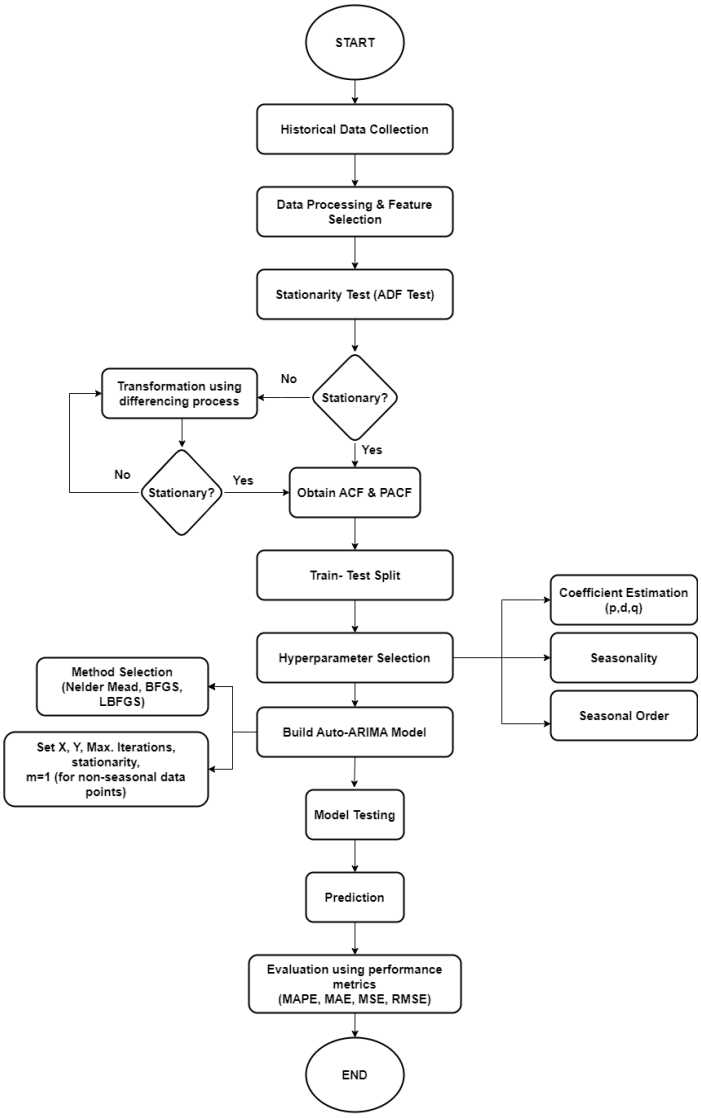


Fig. 2. ARIMA model for wind power forecasting

To begin with, the time series data's stationarity was examined by checking the run order plot and Auto-Correlation function (ACF) plots of the data points. Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots can be used to examine patterns in time series data and the constant variance assumption. Augmented Dickey-Fuller (ADF) Test was also performed to quickly determine the stationarity. In addition, the ACF and PACF plots also aid in determining the AR and MA terms. The values for (p, d, q) were also determined. Fig. 3 and Fig. 4 show the ACF and PACF for the collected daily wind power generation data, respectively.

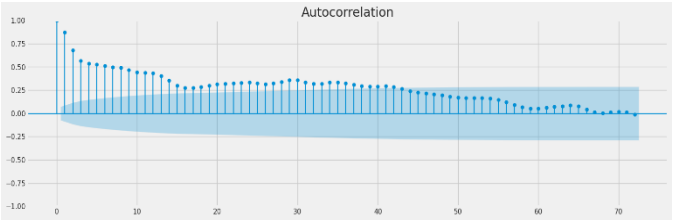


Fig. 3. Autocorrelation functions for wind energy generation data

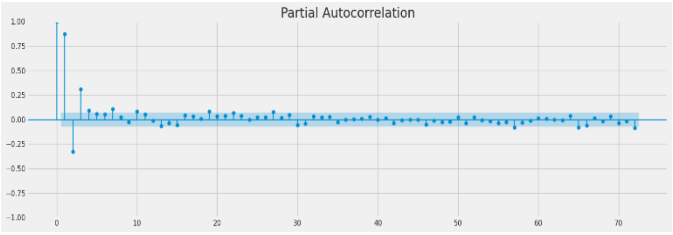


Fig. 4. Partial Autocorrelation functions for wind power generation data

Starting with a visual representation of the time series, several parameter values are then tried intuitively and changed repeatedly until a suitable fit is found. For the ARIMA model, the estimation can be done using an auto-correlation plot for the 'p' parameter, while the partial auto-correlation plot can be used for estimating the 'q' parameter. The Augmented Dickey-Fuller (ADF) test can be used to determine the stationarity of the model and estimate the value for the 'd' parameter.

C. Optimisation Methods

a) Nelder-Mead method:

The NM technique (Nelder-Mead) is popularly used in tackling the unconstrained optimisation problems. It is a direct search approach that does not need the computation of gradients or Hessians.

$$\min f(x), \qquad (7)$$

Where $f: \mathbb{R}^n \to \mathbb{R}$ is called the objective function and n is the dimension. [22]

The NM approach begins with a guess of the parameter values and iteratively seeks the optimal values that minimise the goal function. The approach iteratively constructs an ordered sequence of simplices and attains an optimal point of (7).

At every iteration, the vertices are ordered in line with the values of the objective function.

$$f(x_1) \leq f(x_2) \leq \cdots \leq f(x_{n+1}) \qquad (8)$$

Let $\overline{x}$ be the centroid of the n best vertices, then

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (9)$$

The NM method is easy to construct and is generally used as a foundation for more complicated optimisation methods. The best results are obtained while using the NM method for optimization.

b) Broyden Fletcher Goldfarb Shanno (BFGS) technique:

BFGS (Broyden Fletcher Goldfarb Shanno) is a quasi-Newton approach that applies the knowledge from past iterations to approximate the Hessian matrix. The BFGS technique is a gradient-based optimisation approach that iteratively updates parameter estimates in the negative gradient direction until convergence.

The optimisation problem is equation (4), where f is a differential scalar function and x a vector in $\mathbb{R}^n$. The equivalent solution of Newton equation provides the search direction p_h at h:

$$B_h p_h = -\nabla f(x_h) \qquad (10)$$

Where B_h is approximated to the Hessian matrix at x_h and $\nabla f(x_h)$ is the gradient evaluated at x_h. Then, by minimising $f(x_h + \gamma p_h)$ over the scalar $\gamma > 0$, the next point x_{h+1} is found using a line search in the direction p_h.

The BFGS approach is efficient in terms of computing and can handle a high number of parameters.

c) Limited-Memory BFGS (L-BFGS) method:

The L-BFGS (Low-Memory BFGS) method is a memory-efficient variant of the BFGS method that stores information from earlier iterations in a restricted amount of memory.

Similar to the other two mentioned algorithm, this method also aims to solve minimise the function, $f: \mathbb{R}^n \to \mathbb{R}$. L-BFGS just saves a small number of vectors that implicitly indicate the approximation, whereas BFGS saves a dense $n \times n$times approximation to the inverse Hessian ($n =$ the number of variables).

$$g_h = \nabla f(x_h), \qquad (11)$$

Where f being the function to be minimised.

The L-BFGS method is very beneficial for large-scale optimization problems with many parameters.

D. Performance Metrics

For evaluating the accuracy of the method, Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are employed.

Mean Absolute Error (MAE):

The mean absolute error is the mean of the absolute difference between the actual and predicted numbers. It calculates the mean of the residuals in the dataset. Mathematically, MAE is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\, y_i - \hat{y}_i \,| \qquad (12)$$

Where,

y_i - the actual value,
$\hat{y}_i$ - the predicted value, and
n - the sample size

Mean Absolute Percentage Error (MAPE):

The mean absolute percentage error is defined as the difference between the actual and anticipated values divided

by the actual value. This number is assigned an absolute percentage and averaged throughout the dataset. The performance of the model improves with decreasing MAPE.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} .100\% \quad (13)$$

Mean Squared Error (MSE):

The average of the squared difference between the data set's original and expected values is the Mean Squared Error. It computes the residual variance.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (14)$$

Root Mean Squared Error (RMSE):

RMSE is determined by taking the square root of MSE. It examines the deviation from the actual value and calculates the average magnitude of the errors. RMSE value with zero implies that the model has a perfect fit.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (15)$$

III. EXPERIMENT & EVALUATION

In this research work, the authors have proposed a time-series forecasting model to forecast wind power generation using the ARIMA model. The experiments were conducted on a machine running Windows 11 with an Intel Core i5-10210U CPU, 8 GB RAM, and a 64-bit operating system. Python 3.10 was used throughout for data processing and to develop the ARIMA model.

A. Data Collection

The authors have focused their research on the potential region of Gujarat, more precisely in the Kutch region (Latitude: 23.7337° N, Longitude: 69.8597° E) which according to the data ranks first on the list for wind power potential with 142.56 GW (at 120 M AGL) [4].

The historical data for wind power generation has been gathered from the Central Electricity Authority (CEA), Govt. of India [23], and weather data from the timeanddata website [24]. The data has been recorded daily for more than two years, from 1 January 2021 to 22 January 2023. Fig. 5 represents the wind power generation time series data.

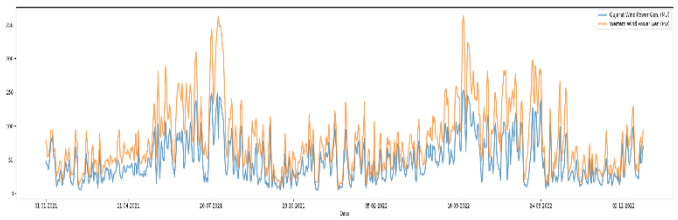


Fig. 5. Power Generation: Western Region of India vs Gujarat

B. Data Pre-processing

In the data processing part, feature extraction was performed and important features for the model training are filtered out. Weather factors like wind speed, wind direction, air density, temperature, and humidity and the wind power generation records were filtered and analysed and checked for seasonality and trends. In the case of NaN and/or infinite values, they were checked and removed from the dataset.

The data collected was analysed and plotted using Python libraries like Plotly and Matplotlib to analyze the trends. It was observed that the general pattern for the wind direction was mostly in the South-West direction followed by North East. Fig. 7 represents the windrose graph for the wind direction. To analyze the correlation between different parameters, a Seaborn heatmap was generated as shown in Fig. 6.

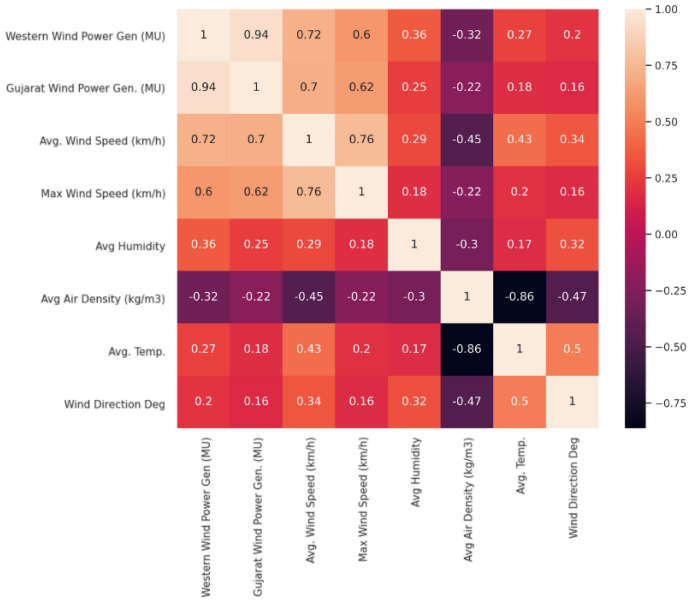


Fig. 6. Correlation Heatmap between the different parameters

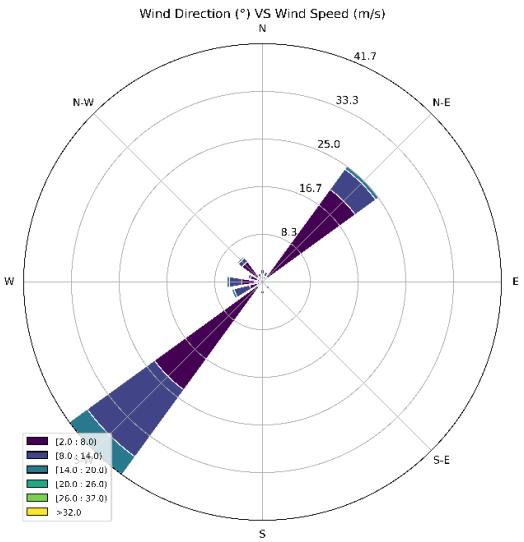


Fig. 7. Windrose plot for Wind Direction in Kutch Region, Gujarat

Boxplots were generated to understand the seasonality patterns for different parameters. The wind power generation for the Gujarat region was observed to be greater between May and August when the demand was higher during the summers, illustrated in Fig. 8.

The seasonality plot for the wind speed showed a similar pattern which depicted a strong correlation between wind power and wind speed of the region. In contrast, the air density plot revealed an opposite trend to the wind speed. Fig. 9 depicts the Annual Seasonality in the Average Wind Speed of the region in kilometres per hour (kmph), while Fig.10 shows the Annual Seasonality of the Average Air Density.

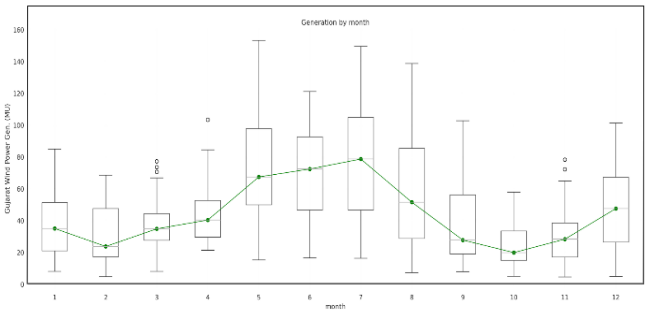


Fig. 8. Annual Seasonality Boxplot: Wind Power Generation

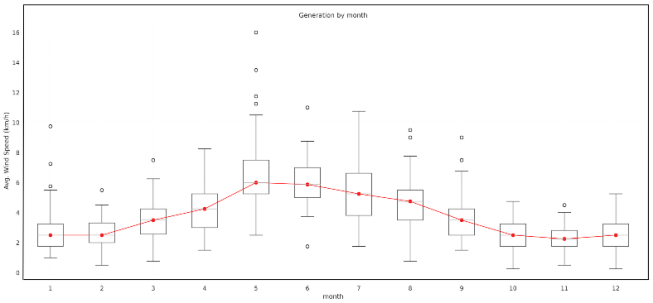


Fig. 9. Annual Seasonality Boxplot: Average Wind Speed (kmph)

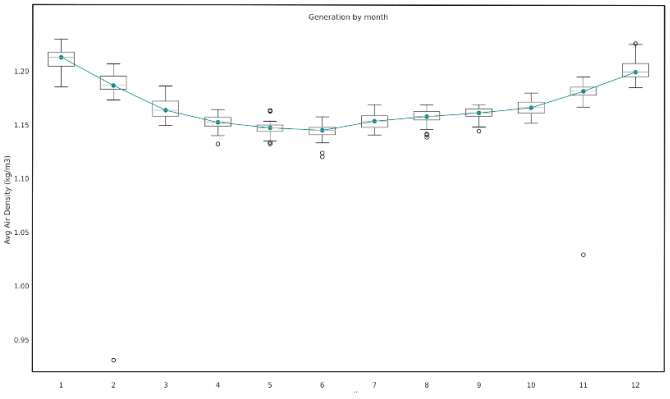


Fig. 10. Annual Seasonality Boxplot: Air Density Generation

C. Evaluation

Four model evaluation criteria were used in this study: the mean absolute error (MAE), which indicates the size of predictive errors, the mean squared error (MSE), which calculates the square of the difference between actual and predicted values, the root mean square error (RMSE), which indicates the severity of discrepancies between actual values and forecasts, and the mean absolute percent error (MAPE), which can avoid the problem that errors cancel each other out and accurately reflect actual data.

The dataset was at first separated into 95% and 5% train and test samples, respectively. Subsequently, different train and test sizes in different ratios were used – 95:5, 96.5:3.5, 97:3, 97.5:2.5, and 98:2, to find the best scores. Auto-ARIMA was applied to select best settings for p, d, and q parameters. The best parameters were obtained at ARIMA (1,2,2) and thereafter the error calculated was optimised and hyperparameter tuning was performed for further results. Auto-ARIMA is the automated version of ARIMA which automatically searches for the best combination of parameters depending on the AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion).

D. Results

Further computations were performed using the ARIMA (1,2,2) model. The model and its performance are assessed in this study using the performance measures MSE, MAE, MAPE, and RMSE. Optimisation algorithms like Nelder-Mead Method, BFGS and L-BFGS methods were adopted to improve performance of the method. The results obtained from model optimisation using NM method are depicted in Table II. Similarly, Table III and Table IV illustrates the results obtained using the BFGS and L-BFGS method, respectively. The authors have considered two different datasets, the Test dataset and Predict dataset, for testing to obtain the best results and verify the applicability of the model and whether it gives desired results with a different dataset. The Test data has been taken between the date range: 05/12/2022 – 26/12/2022 while the Predict Dataset range from 27/12/2022 to 22/01/2023. Table represents the values for the Test Dataset and Table V depicts the values for the Predict Dataset.

In comparison to the other works on wind power forecasting, cited in literature, ARIMA method accurately predicts the forecast for 22 days with 5.87 MAPE on Predict dataset using the NM method (training size of 97% of the main dataset is considered in this). Notably, the 11.50 MAPE was recorded in the test dataset while considering the training size of 95% of the main dataset. For the NM method, 6.24 MAE, 54.88 MSE and 7.4 RMSE were obtained on the test dataset, best performed on training size of 95%, while on predict data (training size: 97%), 3.82 MAE, 20.71 MSE and 4.55 RMSE were obtained, respectively. For the BFGS method, the best results were obtained on a training size of 96.5% of the main dataset on ARIMA (4,2,5) and for the LBFGS method, the best results were obtained on a training size of 97.5% on ARIMA (5,2,0).

Fig. 11 shows the plot between the actual power generation value versus the forecasted values, all values are in Mega Units (MU). It is evident that the model can fairly accurately trace the actual value. Based on the performance solely and reducing hardware limitations, ARIMA can produce benchmark results and is still a popular model for time series forecasting problems. Data quality and availability still can be considered a major drawback for working on projects like these since time series forecasting is crucially based on historical data.

Table I: Test Dataset Results using the NM method

Date	Actual Generation (MU)	Predicted (MU)	Absolute Error	Absolute Error Percentage	Squared Error	Root of Squared Error	Accuracy
05-12-2022	78.78	78.198046	0.581954	0.738707794	0.338670458	0.581954	99.26129221
06-12-2022	101.93	103.847281	1.917281	1.880978122	3.675966433	1.917281	98.11902188
07-12-2022	89.27	91.487016	2.217016	2.483495015	4.915159944	2.217016	97.51650498
08-12-2022	68.34	61.919892	6.420108	9.394363477	41.21778673	6.420108	90.60563652
09-12-2022	73.85	53.134296	20.715704	28.05105484	429.1403922	20.715704	71.94894516
10-12-2022	78.5	64.813535	13.686465	17.43498726	187.3193242	13.686465	82.56501274
11-12-2022	88.76	74.448076	14.311924	16.12429473	204.8311686	14.311924	83.87570527
12-12-2022	96.42	87.796981	8.623019	8.943185024	74.35645667	8.623019	91.05681498
13-12-2022	125.64	112.874385	12.765615	10.16047039	162.9609263	12.765615	89.83952961
14-12-2022	128.56	118.706496	9.853504	7.664517735	97.09154108	9.853504	92.33548227
15-12-2022	75.9	76.940114	1.040114	1.370374177	1.081837133	1.040114	98.62962582
16-12-2022	41.61	42.018831	0.408831	0.982530642	0.167142787	0.408831	99.01746936
17-12-2022	36.11	32.961906	3.148094	8.718067017	9.910495833	3.148094	91.28193298
18-12-2022	42.81	31.090917	11.719083	27.3746391	137.3369064	11.719083	72.6253609
19-12-2022	34.32	30.118811	4.201189	12.24122669	17.64998901	4.201189	87.75877331
20-12-2022	30.85	27.169525	3.680475	11.9302269	13.54589623	3.680475	88.0697731
21-12-2022	54.67	53.524987	1.145013	2.094408268	1.31105477	1.145013	97.90559173
22-12-2022	76.38	85.966734	9.586734	12.55136685	91.90546879	9.586734	87.44863315
23-12-2022	82.9	55.358146	27.541854	33.22298432	758.5537218	27.541854	66.77701568
24-12-2022	58.96	57.643002	1.316998	2.233714383	1.734483732	1.316998	97.76628562
25-12-2022	85.37	76.516726	8.853274	10.37047441	78.38046052	8.853274	89.62952559
26-12-2022	93.09	83.433223	9.656777	10.37359222	93.25334203	9.656777	89.62640778

Table II: Comparison of Results obtained using NM optimization algorithm

Train Size	Test Size	Model	Test Dataset				Predict Dataset			
			MAPE	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE
95.0%	**5.0%**	(1,2,2)	11.507257	6.241513	54.888966	7.408709	9.676041	6.030805	53.226633	7.295658
96.5%	**3.5%**	(1,2,2)	12.531718	7.353864	102.438129	10.121172	7.019101	4.421529	27.950861	5.286857
97.0%	**3.0%**	(1,2,2)	13.167187	7.881410	109.576279	10.467869	5.872073	3.821285	20.717587	4.551658
97.5%	**2.5%**	(1,2,2)	15.680275	9.187369	138.810791	11.781799	6.467925	4.325951	31.246006	5.589813
98.0%	**2.0%**	(5,2,3)	31.921045	29.494326	1052.903111	32.448469	75.163419	47.429191	2599.856705	50.988790

Table III: Comparison of Results obtained using BFGS optimization algorithm

Train Size	Test Size	Model	Test Dataset				Predict Dataset			
			MAPE	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE
95.0%	**5.0%**	(3,2,4)	12.551753	6.840859	62.8912372	7.9303996	11.86181	6.842874	62.46632251	7.903564
96.5%	**3.5%**	(4,2,5)	10.975716	6.959213	69.349459	8.327632	7.893842	4.789487	30.217835	5.497075
97.0%	**3.0%**	(5,2,3)	16.505746	9.558660	156.825551	12.523001	8.463977	5.896083	52.880493	7.271897
97.5%	**2.5%**	(3,2,4)	20.589382	20.589382	182.883915	13.523458	9.709613	6.481890	59.573613	7.718394
98.0%	**2.0%**	(2,2,2)	19.851561	14.127983	264.860741	16.274543	33.084559	19.023742	396.311463	19.907573

Table IV: Comparison of Results obtained using L-BFGS optimization algorithm

Train Size	Test Size	Model	Test Dataset				Predict Dataset			
			MAPE	MAE	MSE	RMSE	MAPE	MAE	MSE	RMSE
95.0%	**5.0%**	(5,2,0)	19.211025	12.598856	213.697871	14.618409	21.636511	13.420372	215.640635	14.684708
96.5%	**3.5%**	(5,2,0)	20.953278	17.151410	426.543925	20.652940	33.994366	21.852226	606.781865	24.632943
97.0%	**3.0%**	(5,2,0)	26.580088	12.823718	241.466078	15.539179	13.859349	9.178889	114.048131	10.679332
97.5%	**2.5%**	(5,2,0)	18.227243	10.476437	163.882306	12.801652	7.405681	4.983938	39.750960	6.304836
98.0%	**2.0%**	(5,2,0)	32.549829	30.541192	1128.614857	33.594864	78.381494	49.668267	2849.454744	53.380284

Table V: Predict Dataset Results using the NM method

Date	Actual Generation (MU)	Predicted (MU)	Absolute Error	Absolute Error Percentage	Squared Error	Root of Squared Error	Accuracy
27-12-2022	78.89	80.410519	1.520519	1.927391304	2.311978029	1.520519	98.07261
28-12-2022	50.58	53.875499	3.295499	6.515419138	10.86031366	3.295499	93.48458
29-12-2022	19.9	20.798685	0.898685	4.516005025	0.807634729	0.898685	95.48399
30-12-2022	66.69	74.077251	7.387251	11.07699955	54.57147734	7.387251	88.923
31-12-2022	98.17	105.457173	7.287173	7.423014159	53.10289033	7.287173	92.57699
01-01-2023	53.27	54.959916	1.689916	3.172359677	2.855816087	1.689916	96.82764
02-01-2023	32.28	28.057056	4.222944	13.08223048	17.83325603	4.222944	86.91777
03-01-2023	100.99	101.891266	0.901266	0.892430934	0.812280403	0.901266	99.10757
04-01-2023	159.44	157.544743	1.895257	1.188696061	3.591999096	1.895257	98.8113
05-01-2023	157.7	152.729402	4.970598	3.151932784	24.70684448	4.970598	96.84807
06-01-2023	117.25	109.599456	7.650544	6.524984222	58.5308235	7.650544	93.47502
07-01-2023	78.87	73.789484	5.080516	6.441633067	25.81164283	5.080516	93.55837
08-01-2023	43.16	40.037467	3.122533	7.234784523	9.750212336	3.122533	92.76522
09-01-2023	33.16	29.578985	3.581015	10.79920084	12.82366843	3.581015	89.2008
10-01-2023	40.12	37.16056	2.95944	7.376470588	8.758285114	2.95944	92.62353
11-01-2023	59.54	63.01681	3.47681	5.839452469	12.08820778	3.47681	94.16055
12-01-2023	99.09	98.748089	0.341911	0.345050964	0.116903132	0.341911	99.65495
13-01-2023	51.42	44.52908	6.89092	13.40124465	47.48477845	6.89092	86.59876
14-01-2023	72.35	75.796122	3.446122	4.763126469	11.87575684	3.446122	95.23687
15-01-2023	76.56	81.497256	4.937256	6.448871473	24.37649681	4.937256	93.55113
16-01-2023	48.44	47.460251	0.979749	2.02260322	0.959908103	0.979749	97.9774
17-01-2023	79.33	79.415832	0.085832	0.108196143	0.007367132	0.085832	99.8918
18-01-2023	76.4	71.439298	4.960702	6.493065445	24.60856433	4.960702	93.50693
19-01-2023	72.88	62.487742	10.392258	14.25940999	107.9990263	10.392258	85.74059
20-01-2023	83.28	79.727137	3.552863	4.266165946	12.6228355	3.552863	95.73383
21-01-2023	87.32	90.48525	3.16525	3.624885479	10.01880756	3.16525	96.37511
22-01-2023	79.32	83.801858	4.481858	5.650350479	20.08705113	4.481858	94.34965

Fig. 11. Wind Power Generation: Actual Values vs Predicted Values (in MU)

IV. CONCLUSION

The authors in this study proposed a statistical model, ARIMA for wind power forecasting in Gujarat, depending on previous past load data and weather data. Modern energy systems require forecasting to operate loads properly, minimize risks and improve system performance. Statistical models, effectively ARIMA, have proved to be efficient prediction methods. The authors of this paper found the challenges with current forecasting models and poorly tuned machine learning models in the application. Forecasting wind power generation is complexly dependent on natural factors like weather factors and technical factors, like the height of windmills, wind turbines, etc. Given the erratic pattern of weather and an added level of uncertainty, the authors attempted to propose a plausible approach for forecasting over the long term that supports both strategic and tactical decisions. With fined tuned parameters, the authors were able to produce accurate results in comparison to previous historic data. With the proposed method, it is possible to precisely forecast the generation for each day.

V. ACKNOWLEDGEMENT

The authors want to claim that there exists no known opposing interest or close ties which might affect the outcomes of the research presented in this study.

VI. REFERENCES

[1] Ministry of New and Renewable Energy, Government of India, "Physical Progress (Achievements)," 2020.

[2] S. Saji, N. Kuldeep and A. Tyagi, "A Second Wind for India's Wind Energy Sector: Pathways to achieve 60 GW," Council on Energy, Envirnoment and Water, New Delhi, 2019.

[3] "National Power Portal," National Informatics Centre, Ministry of Electronics & Information Technology, Government of India, [Online]. Available: https://npp.gov.in/dashBoard/cp-map-dashboard. [Accessed 18 April 2023].

[4] A. Verma, "Saur Energy International," 10 June 2020. [Online]. Available: https://www.saurenergy.com/solar-energy-blog/top-10-states-india-wind-energy. [Accessed April 2023].

[5] X. Chen, X. Zhang, M. Dong, L. Huang, Y. Guo and S. He, "Deep Learning-Based Prediction of Wind Power for Multi-turbines in a Wind Farm," *Frontiers in Energy Research,* vol. 9, 2021.

[6] D. Sharma and R. Thakur, "A Study on Recent Trends for Load Forecasting with Artificial Intelligence," *Turkish Online Journal of Qualitative Inquiry,* vol. 12, no. 6, 2021.

[7] H. Hahn, S. Meyer-Nieberg and S. Pickl, "Electric load forecasting methods: Tools for decision making," *European Journal of Operational Research, Elsevier,* vol. 199, no. 3, 2009.

[8] E. Kyriades and M. Polycarpou, "Short Term Electric Load Forecasting: A Tutorial," *Trends in Neural Computation,* vol. 35.

[9] K. Gajowniczek and T. Ząbkowski, "Short Term Electricity Forecasting Using Individual Smart Meter Data," *Procedia Computer Science,* vol. 35, pp. 589-597, 2014.

[10] Y. Nagaraja, T. Devaraju, M. V. Kumar and S. Madichetty, "A survey on wind energy, load and price forecasting: (Forecasting methods)," *International Conference on Electrical Electronics and Optimization Techniques (ICEEOT),* pp. 783-788, 2016.

[11] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews,* vol. 13, no. 4, pp. 915-920, 2009.

[12] A.-N. Buturache and S. Stancu, "Wind Energy Prediction Using Machine Learning," *Low Carbon Economy,* vol. 12, no. 1, 2021.

[13] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renewable and Sustainable Energy Reviews,* vol. 31(c), pp. 762-777, 2014.

[14] E. Grigonytė and E. Butkevičiūtė, "Short-term wind speed forecasting using ARIMA model," *Energetika,* vol. 62, 2016.

[15] E. Yatiyana, S. Rajakaruna and A. Ghosh, "Wind speed and direction forecasting for wind power generation using ARIMA model," in *Australasian Universities Power Engineering Conference (AUPEC),* 2017.

[16] I. Tyass, A. Bellat, A. Raihani, K. Mansouri and T. Khalili, "Wind Speed Prediction Based on Seasonal ARIMA model," *The International Conference on Energy and Green Computing,* vol. 336, 2022.

[17] B.-M. Hodge, A. Zeiler, D. Brooks, G. Blau, J. Pekny and G. Reklatis, "Improved Wind Power Forecasting with ARIMA Models," *Computer Aided Chemical Engineering,* vol. 29, pp. 1789-1793, 2011.

[18] J. Shi, X. Qu and S. Zeng, "Short-Term Wind Power Generation Forecasting: Direct Versus Indirect Arima-Based Approaches," *International Journal of Green Energy,* vol. 8, no. 1, pp. 100-112, 2011.

[19] P. Du, J. Wang, W. Yang and T. Niu, "A novel hybrid model for short-term wind power forecasting," *Applied Soft Computing,* vol. 80, 2019.

[20] C. Draxl, A. Purkayastha and Z. Parker, "National Renewable Energy Laboratory (NREL)," July 2014. [Online]. Available: https://www.nrel.gov. [Accessed April 2023].

[21] G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control, Wiley, 1970.

[22] F. Gao and L. Han, "Implementing the Nelder-Mead simplex algorithm with adaptive parameters," *Computational Optimization and Applications,* vol. 51, pp. 259-277, 2012.

[23] Ministry of Power, Government of India, "Central Electricity Authority of India," [Online]. Available: https://cea.nic.in. [Accessed April 2023].

[24] S. Thorsen, "timeanddate," 1998. [Online]. Available: https://www.timeanddate.com. [Accessed February 2023].

VII. BIOGRAPHIES

Sulagna Mahata was born in Patna, India, in 2001. She graduated in Electrical Engineering in 2023 from BIT Sindri, Dhanbad, India. She was also research intern at McMaster University, Hamilton in Canada. Her research interests include time series forecasting, machine learning and artificial intelligence.

Piyush Harsh was born in Dhanbad, India, in 2001. He graduated in Electrical Engineering in 2023 from BIT Sindri, Dhanbad, India. He also interned at Qatar University and IIT Guwahati. He developed an interest in machine learning and artificial intelligence. His research interests include Time-Series Forecasting, Natural Language Processing, Machine Learning and Artificial Intelligence.

Vineet Shekher graduated from the Department of Instrumentation Engineering of Gangamai College of Engineering affiliated to North Maharashtra University, Jalgoan in 1999. He received his post-graduation in Electrical Engineering with specialization in Control and Instrumentation from Delhi College of Engineering affiliated to Delhi University, Delhi, India, in 2009. He was conferred with Doctor of Philosophy in 2016 in the field of Electrical Engineering specialized in the area of fractional system in BIT, Sindri, Dhanbad affiliated to Vinoba Bhave University, Hazaribagh, Jharkhand, India. He worked as Lecturer and Assistant Professor in Various Engineering Institutes. Since March 2021 he is working as an associate professor in the Department of Electrical Engineering, BIT, Sindri, Dhanbad, Jharkhand, India. He has published more than 10 research papers in various SCI journals holding high impact factors. He has also published research articles in national journals and international conference proceedings. His research interests include stochastic differential equations, dynamical systems, fuzzy neural networks and complex networks.

Pankaj Rai graduated in the Department of Electrical Engineering from MIT, Muzaffarpur affiliated to Bihar University, Muzaffarpur in 1988. He received his post-graduation in Electrical Engineering with specialization in Control System from BIT Sindri affiliated to Vinoba Bhave University, Hazaribagh, Jharkhand, India, in 2002. He was conferred with Doctor of Philosophy in 2009in the field of Electrical Engineering specialized in the area of Soft Computing in BIT, Sindri, Dhanbad affiliated to Vinoba Bhave University, Hazaribag, Jharkhand, India. He is working as Professor in the Department of Electrical Engineering, BIT, Sindri, Dhanbad, Jharkhand, India. He has published research papers in various SCI journals holding high impact factors. He has also published research articles in national journals and international conference proceedings. His research interests include stochastic differential equations, dynamical systems, fuzzy, neural networks and complex networks.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

A SURVEY PAPER ON INTELLIGENT INTRUSION DETECTION SYSTEM BASED ON DEEP LEARNING APPROACH

Prateek Shrivastava and Rajesh Kumar Yadav
Computer Science Engineering, Delhi Technological University

Abstract. In order to achieve necessary security assurance, intrusion detection systems (IDSs) play a vital role in all networks & information systems worldwide. One method for reducing malicious attacks is IDS. As attackers constantly adapt their attack strategies & discover new attack vectors, IDS must also advance by implementing more complex detection mechanisms. New studies in the deep learning sector, including intrusion detection, are now available as a result of the enormous rise in data & major improvements in computer hardware technology. A branch of Machine Learning (ML) techniques based on learning data representations is called deep learning. This study begins by providing a thorough overview of various deep-learning techniques used in IDSs. primary works that have been described in deep learning studies are then summarised, followed by a presentation about a deep learning classification scheme. We have used this method to present a taxonomy review of deep architectures & algorithms that are available in these works & to categorize those algorithms into three classes: discriminative, hybrid, & generative. Then, selected deep learning applications are examined in a variety of intrusion detection sectors. Finally, prevalent dataset & framework kinds are covered.

KEYWORDS: Intrusion Detection Systems, Recurrent Neural Network, Deep Learning, Deep Neural Network.

1. INTRODUCTION

For a very long time, computer & network system security has been focus about research. Information protection is a highly essential & critical issue that cannot be neglected, according towards all businesses working in field about information technology. three fundamental tenets on which any safe system is built must be attained (confidentiality, integrity, & availability). Intrusion detection is described as "the process about monitoring events occurring in a computer system otherwise network & analysing them for signs about intrusions, defined as attempts towards compromise confidentiality, integrity, otherwise availability about a computer otherwise network" through National Institute about Standards & Technology. [1],[2]. Systems & networks about official & unofficial organisations, e-commerce, & even individuals worldwide are subject towards new kinds about cyberattacks every day. In order towards interrupt operation about these systems, which totally rely on this information, various attempts are made towards either gain specific information otherwise delete information itself. One answer towards these issues & advancements is intrusion detection systems (IDS) [3]. IDSs have been able towards identify many forms about computer system utilisation & harmful network connections, whereas traditional firewall is unable towards carry out this operation. IDSs' working premise is based on distinction between legitimate user & intruder usage.

IDSs are often parted into two classes. Impromptu, in light of their procedures for abuse (signature) recognition, and specially appointed [5]. reason for abnormality identification is towards decide if varieties from perceived utilization examples can be accounted for as interruptions. Abuse identification utilizes instances of notable goes after in any case framework shortcomings towards learn whether interruptions have happened [6]. Ongoing years have seen a huge expansion being used of deep learning in research and a great many applications, including picture classification, information mining, data security, including interruption location, and extraction and IDSs

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

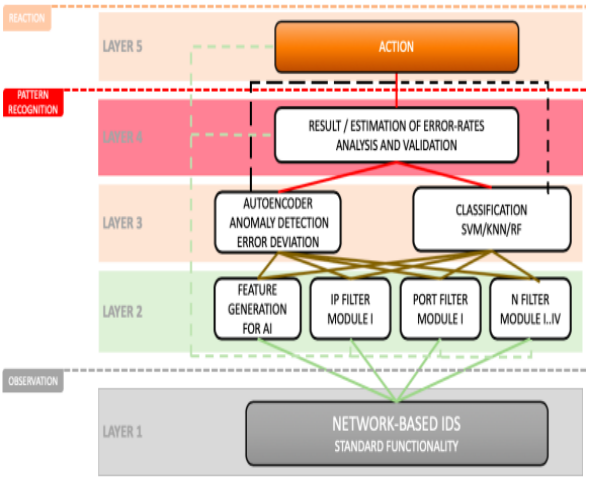


Fig.1: Intelligent IDS architecture

are often parted into two classes. Impromptu, in light of their procedures for abuse (signature) recognition, and specially appointed [5]. reason for abnormality identification is towards decide if varieties from perceived utilization examples can be accounted for as interruptions. Abuse identification utilizes instances of notable goes after in any case framework shortcomings towards learn whether interruptions have happened [6]. Ongoing years have seen a huge expansion being used of deep learning in research and a great many applications, including picture classification, information mining, data security, including interruption location, and extraction and examination of video information. [7]. A subset of AI (ML) approaches called deep learning involves various data handling layers in progressive designs towards track down designs and learn new highlights in any case portrayals [8]. Deep learning is presently an extremely unmistakable and effective examination pattern in ML people group due towards its extraordinary outcome in these fields [9]. In this review, we sum up greater part of ongoing examinations that have utilized deep learning techniques with interruption recognition frameworks.

Networks are progressively impacting day to day existence, subsequently understanding digital protection is fundamental. most famous network safety devices are interruption location frameworks, antivirus programming, and firewalls (IDSs). These techniques monitor networks against dangers from both inside and outside organization. For instance, an IDS is a kind of identification framework that screens programming and equipment designs working on an organization and helps towards guard network safety. In 1980, first interruption recognition framework was proposed [1]. From that point forward, a ton of data on reliable IDS items has surfaced. Numerous IDSs, be that as it may, go on towards have a high deception rate, conveying various signs for low-danger conditions, tiring security examiners, and here and there permitting extreme assaults towards escape inconspicuous. In this manner, numerous researchers definitely stand out towards making of IDSs with higher discovery rates and decreased deception rates. One more issue with current IDSs is their disappointment towards recognize unreported attacks. A constant deluge of new assault types and goes after results from fast organization climate change. It is fundamental towards foster IDSs that can identify obscure dangers. towards tackle previously mentioned issues, specialists are focusing on creating IDSs utilizing AI methods. AI, a sort of man-made brainpower, can naturally remove pertinent information from large data sets. [2]. While adequate preparation information is accessible and AI models have enough generalizability towards perceive assault varieties and interesting attacks, AI based IDSs might accomplish extraordinary discovery levels. Also, in light of the fact that they don't considerably depend on space ability, AI based IDSs are straightforward towards make and develop. Deep learning is an AI strategy that can give first class results. When looked at towards conventional AI strategies, deep learning approaches are more adroit at overseeing monstrous measures of information [3]. Since they work start to finish and can consequently gain highlight portrayals from crude information, deep learning methods are especially helpful. Deep learning is recognized by its deep construction, which incorporates various covered levels.

2. LITERATURE REVIEW

2.1 NIST special publication on intrusion detection systems:

IDSs are programming in any case equipment gadgets that robotize cycle of checking and dissecting PC in any case network movement for indications of safety issues. Due towards expansion

in recurrence and seriousness of organization assaults over beyond couple of years, interruption recognition frameworks are currently an essential piece of safety foundation for greater part of organizations. For the individuals who need towards understand what security objectives these components support, how towards pick and design interruption discovery frameworks for their specific framework and organization conditions, how towards oversee yield about interruption location frameworks, and how towards incorporate interruption identification capabilities among different parts of hierarchical security foundation, this guide was composed as a presentation towards interruption recognition. References that give peruser concentrated in any case top to bottom data on unambiguous interruption location issues references towards extra data sources are likewise advertised.

2.2 Intrusion detection: A survey:

The outlook for network security has changed due towards increasing spread about computer networks. A state that makes information easily accessible makes computer networks susceptible towards various hacker attacks. There are numerous & potentially catastrophic threats towards networks. Researchers have so far created intrusion detection systems (IDS) that can recognize attacks in a variety about situations. There are countless techniques that can be used towards identify misuse & anomalies. Since different types about ecosystems are best served through different approaches, many about technologies presented are complementary towards one another. In order towards survey & categorise intrusion detection systems, this study proposes a taxonomy for doing so. detection theory & a few operational components about intrusion detection make up taxonomy.

2.3 Survey on Intrusion Detection System using Machine Learning Techniques:

In current world, nearly everybody approaches towards a PC, and organization based innovation is quickly developing. Network security has developed towards be a significant, if not irreplaceable, part of PC frameworks. objective of an interruption recognition framework (IDS) is towards perceive framework assaults and separate normal utilization designs from unusual ones. AI procedures have further developed interruption recognition frameworks and different frameworks utilized towards distinguish interruptions. This study surveys an assortment of machine procedures for interruption discovery frameworks. This study's framework engineering for an interruption recognition framework is additionally portrayed, with point of lessening deception rates and further developing interruption discovery exactness.

2.4 Feature Selection for Intrusion Detection System Using Ant Colony Optimization:

Intrusion detection is a key examination point in network security. As a result of nonlinear nature of interruption endeavors, unusual organization traffic conduct, and numerous factors in issue space, interruption identification frameworks are a troublesome area of exploration. Choosing productive and urgent parts for interruption recognition is an exceptionally fundamental subject in data security. point of this study is towards recognize key components for making an interruption identification framework that is both viable and computationally effective. This study suggests an interruption recognition framework whose highlights are painstakingly chosen utilizing insect province advancement towards further develop execution. recommended technique is simple towards use and has little handling intricacy since it involves a little arrangement of highlights for order. As exhibited by significant test discoveries on KDD Cup 99 and NSL-KDD interruption location benchmark informational collections, new technique beats prior draws near, giving more noteworthy exactness in recognizing interruption endeavors and diminished misleading problem among less elements.

2.5 Design about experiments application, concepts, examples: State about art:

Application areas for statistical tool known as Design about Experiments (DOE) include system, process, and product design, development, and optimization. It is a flexible tool that may be used in a wide range of situations, such as design for comparisons, variable screening, transfer function discovery, optimization, and resilient design. state-of-the-art in DOE use is discussed, along with evolution of DOE over time. Researchers are also given instructions on how towards design, organise, and conduct experiments, as well as how

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

towards assess and interpret results using examples. This article also shows how, over past 20 years, DOE applications have been rapidly growing in both manufacturing and non-manufacturing industries. About half of its applications are in sciences, specifically in domains of computer science, engineering, biochemistry, and medicine.

3. DEEP LEARNING APPROACHES

Networks in deep learning, a subset of AI (ML) in field of man-made brainpower (man-made intelligence), can gain from marked and unlabeled information in both directed and unaided techniques. Deep brain networks in any case deep brain learning are different names for deep learning. Deep Learning is an element of artificial intelligence that recreates how human mind functions as far as how it processes information and makes designs that might be applied towards decision-production [9]. In spite of the fact that there is no single meaning of deep learning, greater part of details stress following characteristics:

- Branch about machine learning. Usually nonlinear models.
- Fits models to data using both supervised & unsupervised methods.
- Models are multi-layered graph structures (networks) (deep).

The majority about research in this area & implemented algorithms in subject about intrusion detection can be broadly divided into three primary groups, which are [10] (Figure 2 provides an illustration of deep learning approaches' classification.)

(1) Imaginative (unsupervised).
(2) Inequitable (supervised).
(3) A deep hybrid architecture.

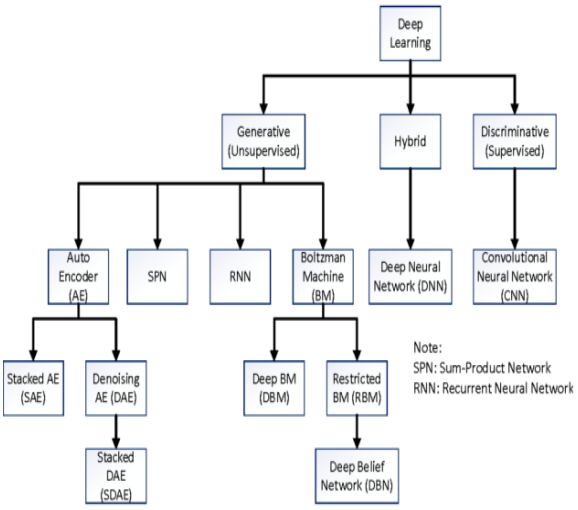


Fig.2: Taxonomy about deep learning methods

4. POPULAR INTRUSION DETECTION DATASETS FOR DEEP LEARNING

For their own examination as well as towards add towards local area vaults, many exploration groups today gather a scope of information sorts. most well-known interruption discovery datasets utilized in DL research are made sense of here.

MIT Lincoln Lab has assembled and dispersed first standard information for assessing PC network IDS under help from "Guard Progressed Exploration Ventures Office" (DARPA) and "Flying corps Exploration Lab" (AFRL). Scientists should remove properties from documents all together towards use them in ML calculations since DARPA informational index is comprised of generally crude records.

The KDDCup 1999 information assortment was used in DARPA IDS assessment program. information comprises of a packed 4 GB tcp dump produced over course of close to seven weeks of organization movement. Every association record, which is around 100 bytes in length, can oblige 5 million associations. There are around 4,900,000 single association vectors in it, and everyone has 41 properties.

The KDDCup 1999 informational collection and NSLKDD informational collection are primarily very comparative (at the end of the day it has 22 examples about assaults in any case ordinary rush hour gridlock, and fields for 41 properties). Figure 3 shows an overall portrayal of these interconnected informational indexes. (NSLKDD, KDD-99, and DARPA). DARPA is a major crude informational collection. size-diminished and

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

duplications eliminated NSLKDD information assortment relates towards KDD-99.

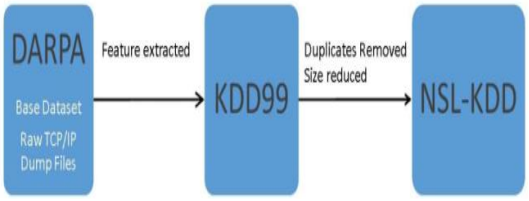


Fig.3: Correlation between main & extracted datasets

5. FRAMEWORKS FOR DEEP LEARNING IMPLEMENTATION

Deep learning architecture combines the implementation about modularized deep learning algorithms among approaches for optimization, dissemination, & infrastructure support. most popular frameworks for implementing deep learning algorithms are briefly introduced in this section.

5.1 Tensor-flow Google has been utilizing Tensor-Flow (TF), the distributed method for training NNs that replaces Dist-Belief, since 2011. Google brain team developed TF, an open-source library for numerical computation. TF operates more quickly since its Python API was used throughout development rather than a C/C++ engine. TF supports CUDA. TensorFlow can be used to create almost any form of network, despite fact that deep networks cannot be configured among hyper-parameters. Additionally, Tensor-Flow includes a C++ interface.

5.2 Theano The ML group about Montreal University created Theano. It is an open-source, cross-platform Python library. multidimensional array mathematical statement is defined, optimised, & evaluated using Theano Python module. High network modelling capability, dynamic code generation, & speed among a variety about GPU support are all provided through Theano. However, Theano offers low-level API & has numerous intricate compilations that are typically time-consuming. Theano, on other hand, has a variety about instructional materials & is still used through a sizable number about academics & developers.

5.3 Keras For implementing deep learning in Python-based Theano & TF, Keras has been created. It enables high-level NN API for swift deep learning algorithm implementation. main selling point about Keras is that it works among Theano & TF, two commonly used deep learning implementation frameworks, & that it can be extended, modularized, & utilised on a user platform using Python. Theano & TF's design makes it easy towards create high-level libraries like Keras that could be used among any about backends. In general, TF & Theano programmes are larger than Keras-equivalent programmes. Keras model is depicted in Figure 4.

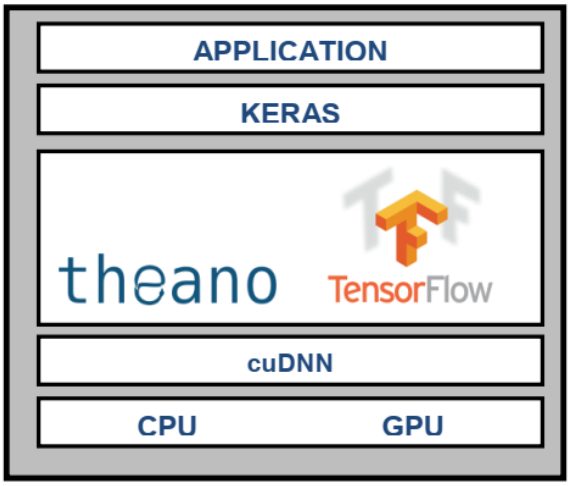


Fig.4: Architecture about Keras

5.4 Torch/PyTorch In view of simple to-utilize, speedy to-learn, and versatile Lua programming language, Light is an open source deep learning system. This structure is a broad ally about ML strategies and is intended for logical calculation. In deep learning structure local area, Py-Light has as of late seen a significant level about prominence and is seen as Tensor-rival. Stream's Py-Light is essentially a port about Light system, which is utilized towards fabricate deep brain organizations and do tensor estimations that are incredibly complicated. A front-end Light incorporation for suitable execution deep learning improvement among critical GPU support has as of late been made at Facebook and is called Py-Light. It ensures a Python front-end that makes it conceivable towards construct dynamic NN. On opposite side, tool compartment has as of late been made accessible, and there isn't any local area support, informative materials, in any case assessment about its adequacy.

6. CONCLUSION

The overview about deep learning & points that majority about definitions emphasis are provided in this essay. We examined most recent articles on deep learning for intrusion detection. We look at a few popular deep learning architectures & highlight some about its applications towards intrusion

detection. More precisely, Generative (unsupervised), Discriminative (supervised), & Hybrid deep architecture classes about deep learning architectures are covered in detail along among their methodologies. These three classes offer a great deal about versatility & have been effective & trustworthy in a variety about challenges for decades. We have examined associated works for each about aforementioned classes & techniques that are used in intrusion detection sector. most well-liked deep learning implementation frameworks and intrusion detection datasets are highlighted in section that follows. Although supervised learning algorithms operate with labelled data, they struggle towards perform well when dealing with large amounts of data since it is difficult towards collect labelled data. In order towards process unlabeled data, unsupervised learning methods are employed. We can also utilise unsupervised learning algorithms towards forecast best results if we are unsure about output data (outputs). Sets about intrusion detection data are crucial for system testing & training. Every dataset has a great number about features, majority about which are superfluous otherwise unimportant. Deep learning techniques are best suited for simplifying complex features otherwise extracting features. If we are unsure about relationship between targeted classification output & raw input data, we may employ deep learning techniques.

In conclusion, it can be claimed that majority about strategies presented have demonstrated ability towards achieve high accuracy levels in a more automatic manner.

7. FUTURE SCOPE

As a viable route for future study employing RNN-based methods towards be implemented in models for accuracy gains, it is advised towards use feature extraction & feature selection as a hybrid strategy towards increase accuracy about intrusion detection.

REFERENCES

[1] R. Bace & P. Mell, "NIST special publication on intrusion detection systems," BOOZ-ALLEN & HAMILTON INC MCLEAN VA, 2001.

[2] A. Lazarevic, V. Kumar, & J. Srivastava, "Intrusion detection: A survey," in Managing Cyber Threats, Springer, 2005, pp. 19–78.

[3] S. K. Wagh, V. K. Pachghare, & S. R. Kolhe, "Survey on intrusion detection system using machine learning techniques," Int. J. Comput. Appl., vol. 78, no. 16, 2013.

[4] W. Stallings, "Cryptography & network security: principles & practice," Pract. (6th Ed., vol. 9, p. 9685, 1998.

[5] M. H. Aghdam & P. Kabiri, "Feature Selection for Intrusion Detection System Using Ant Colony Optimization.," IJ Netw. Secur., vol. 18, no. 3, pp. 420–432, 2016.

[6] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, & W.-Y. Lin, "Intrusion detection through machine learning: A review," Expert Syst. Appl., vol. 36, no. 10, pp. 11994–12000, 2009.

[7] B. Durakovic, "Design about experiments application, concepts, examples: State about art," Period. Eng. Nat. Sci., vol. 5, no. 3, 2017.

[8] S. Pouyanfar et al., "A Survey on Deep Learning: Algorithms, Techniques, & Applications," ACM Comput. Surv., vol. 51, no. 5, p. 92, 2018.

[9] Y. LeCun, Y. Bengio, & G. Hinton, "Deep learning. nature 521 (7553): 436," Google Sch., 2015.

[10] L. Deng & X. Li, "Machine learning paradigms for speech recognition: An overview," IEEE Trans. Audio. Speech. Lang. Processing, vol. 21, no. 5, pp. 1060–1089, 2013.

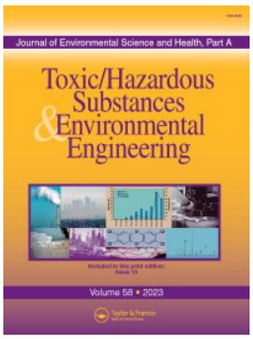
[11] L. Deng & D. Yu, "Deep learning: methods & applications," Found. Trends® Signal Process., vol. 7, no. 3–4, pp. 197–387, 2014.

[12] Y. Bengio, N. Boulanger-Lewandowski, & R. Pascanu, "Advances in optimizing recurrent networks," in 2013 IEEE International Conference on Acoustics, Speech & Signal Processing, 2013, pp. 8624–8628.

[13] E. Aminanto & K. Kim, "Deep learning in intrusion detection system: An overview," in 2016 International Research Conference on Engineering & Technology (2016 IRCET), 2016.

[14] N. Shone, T. N. Ngoc, V. D. Phai, & Q. Shi, "A deep learning approach towards network intrusion detection," IEEE Trans. Emerg. Top. Comput. Intell., vol. 2, no. 1, pp. 41–50, 2018.

[15] I. Goodfellow, Y. Bengio, A. Courville, & Y. Bengio, Deep learning, vol. 1. MIT press Cambridge, 2016.



Taylor & Francis
Taylor & Francis Group

Journal of Environmental Science and Health, Part A
Toxic/Hazardous Substances and Environmental Engineering

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lesa20

Adsorption of anionic dye from aqueous environment using surface-engineered Zn/Cu hydroxy double salt-based material: mechanistic, equilibrium and kinetic studies

Bhamini Pandey, Poonam Singh & Vinod Kumar

To cite this article: Bhamini Pandey, Poonam Singh & Vinod Kumar (2023) Adsorption of anionic dye from aqueous environment using surface-engineered Zn/Cu hydroxy double salt-based material: mechanistic, equilibrium and kinetic studies, Journal of Environmental Science and Health, Part A, 58:10, 869-884, DOI: 10.1080/10934529.2023.2243193

To link to this article: https://doi.org/10.1080/10934529.2023.2243193

Published online: 09 Aug 2023.

Submit your article to this journal

Article views: 25

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lesa20

JOURNAL OF ENVIRONMENTAL SCIENCE AND HEALTH, PART A
2023, VOL. 58, NO. 10, 869–884
https://doi.org/10.1080/10934529.2023.2243193

Taylor & Francis
Taylor & Francis Group

Check for updates

Adsorption of anionic dye from aqueous environment using surface-engineered Zn/Cu hydroxy double salt-based material: mechanistic, equilibrium and kinetic studies

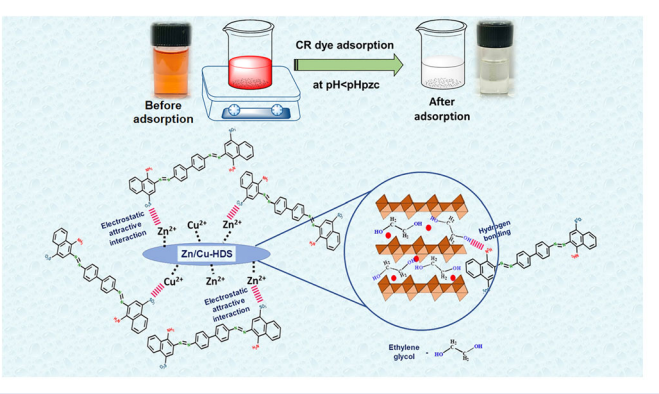
Bhamini Pandey[a], Poonam Singh[a], and Vinod Kumar[b]

[a]Department of Applied Chemistry, Delhi Technological University, Delhi, India; [b]Special Centre for Nanoscience, Jawaharlal Nehru University, Delhi, India

ABSTRACT

Herein, ethylene glycol (EG)-modified Zn/Cu hydroxy double salts (HDS) were synthesized using a facile synthetic approach. The formation of layered structure and presence of EG in the interlayer region were confirmed using PXRD and FTIR techniques. Furthermore, XPS analysis was used to confirm presence of metal ions in synthesized HDS. The surface area and pore size diameter of the HDS was found to be 32.30 m^2 g^{-1} and 2.22 nm, respectively, using BET. The role of HDS was evaluated for its potential application as a sorbent for Congo red (CR) dye uptake. Batch studies were conducted to examine the impact of key variables, i.e., pH, time, adsorbent dosage and dye concentration on adsorption efficiency of HDS. Linear-nonlinear isotherm and kinetic models were employed for detailed analysis of experimental data. Langmuir, Freundlich and Temkin isotherm models were subsequently utilized to fit equilibrium data, among which Langmuir demonstrated to be most accurate. The maximum monolayer adsorption capacity estimated using Langmuir model was computed to be 181.81 mg g^{-1}. The kinetic data follows pseudo-second-order model having good R^2 value (0.999). Additionally, thermodynamic study suggested spontaneous and endothermic nature of adsorption process having reusability up to 5 cycles with removal efficiency more than 85%.

GRAPHICAL ABSTRACT



ARTICLE HISTORY
Received 3 November 2022
Accepted 23 July 2023

KEYWORDS
Hydroxy double salt; ethylene glycol; mesoporous materials; water remediation; monolayer adsorption

Introduction

During the last few years, layered materials having two-dimensional (2D) structures have garnered enormous attention owing to their exceptional physiochemical and optical properties giving rise to potential applications in the field of catalysis, photonics, electronics, drug delivery, energy storage, etc.[1–4] Among these, the materials based on layered double hydroxides (LDHs) are widely explored owing to their distinctive layered structure having a large surface area, presence of different divalent and trivalent ions in the lattice, and high anion exchangeability for numerous organic, polymeric and inorganic species. In general, LDHs are characterized by the formulation $[M^{II}_{1-a}M^{III}_{\ a}(OH)_2]^{a+}$ $A^{x-}_{a/x}\cdot yH_2O$ where M^{II} is the divalent cation (Zn^{2+}, Mn^{2+}, Mg^{2+}, etc.), M^{III} is the trivalent cation (Ni^{3+}, Fe^{3+}, Ga^{3+}, etc.), A^{x-} is the anion in interlayer region of valency x (CO_3^{2-}, CH_3COO^-, SO_4^{2-}, etc.) and a is the molar ratio ($M^{3+}/M^{2+}+M^{3+}$). LDH possesses a brucite-like structure in which there is an isomorphous replacement of a part of M^{II} ion by M^{III} ion resulting in an additional positive charge on the brucite layers, counterbalanced by the existence of anionic species in the interlamellar region.[5,6]

Hydroxy double salts (HDS) are considered another promising class of layered compounds, consisting of two

CONTACT Poonam Singh poonam@dtu.ac.in Department of Applied Chemistry, Delhi Technological University, Delhi, India.
© 2023 Taylor & Francis Group, LLC

types of divalent metal cations (M^{II}). In HDS, one-quarter of octahedral sites are empty, and M^{II} cations are tetrahedrally coordinated just above and below the empty sites, as a consequence, a surplus of positive charge is produced, which is neutralized by charge compensating ions (A^{m-}) in the interlamellar region. The general composition of HDSs is $M^{a}_{1-y}M^{b}_{y}(OH)_{2-x}(A^{m-})_{x/m}.nH_2O$ where M^a and M^b are divalent metal ions and A^{m-} is an interlamellar anion.[7] Subsequently, high surface area, high porosity and anionic exchange capacity of HDS result in efficient sorption of various anionic species.[8] The compositional diversity and intercalation chemistry enables these HDSs to be employed in a broad range of applications including wastewater remediation, drug delivery, catalysis, etc. Kaassis et al. performed the intercalation of various bioactive anions in the Co/Zn and Ni/Zn HDSs and investigated their performance in drug delivery applications.[9] Li et al. prepared another layered material for glucose detection using nanoflowers of Cu HDS.[10] Lu et al. presented the formation of gold nanoclusters that supported Zn HDS for the sensing of silver ions.[11]

Over the past few decades, environmental problems have been aggravated as a result of the release of heavy metals and organic pollutants into water resources. Dyes are one of these pollutants, which are released into water resources from numerous sources including textile, paper, cosmetics, plastic, paint, food, and petrochemical industries. Most dyes, or their metabolites, are toxic, mutagenic, and cancerous in nature.[12,13] The problems relating to dye effluent treatments arise from their differing synthetic formulations and aromatic compositions. They are indeed not eco-friendly either. Congo red (CR), an anionic diazo dye is composed of benzidine-based compounds that are allergic and carcinogenic.[14]

Traditional wastewater treatment processes cannot completely remove contaminants, so advanced technologies are needed to eliminate them from aqueous environments. In addition to enhanced oxidation technologies, constructed wetlands, adsorption, and filtration strategies are employed.[15] Adsorption is the most frequently used environmental remediation technique which is an economical, effective, flexible, and eco-friendly technique. It enables the development of an energy-efficient separation process with affordable procedures without toxic end products.[16,17] Various adsorbents have been utilized for the removal of pollutants in synthetic and real wastewater systems, such as activated carbon, biochar, metal-organic frameworks, layered materials, metal oxide nanoparticles, magnetic nanocomposites, etc.[18–23]

In the past, HDS has been shown to remove several contaminants from the water *via* adsorptive removal.[24] Copper-containing hydroxy double salts (Cu-M HDSs) have been synthesized by Weeramonkhonlert et al. and its catalytic application for the degradation of carcinogenic methyl orange (MO) dye was investigated.[25] The synthesis of various organic anion-modified HDSs reported in the literature led us to consider these materials as sorbents for organic pollutants and particularly the anionic species. Table 1 outlines some of the sorbents identified in the literature for removing the CR dye based on their maximum adsorption capacities.

To the best of our knowledge, a similar lattice has been reported in literature, where EG was incorporated *via* a two-step method in the layered material. The results of the study demonstrated that removal of CR dye was achieved within 70 min of contact time and the maximum adsorption capacity of the adsorbent reported was 91.7 mg g^{-1}.[31] Therefore, an attempt has been made to synthesize Zn/Cu HDS *via* a one-step method with enhanced adsorptive properties for CR dye uptake. In this work, one-pot synthesis of EG-modified Zn/Cu HDS has been carried out. The synthesized lattice is then employed as an adsorbent to eliminate the carcinogenic CR dye. Detailed analysis of the impact of various parameters on the uptake of pollutants was conducted in batch studies. Additionally, the data were also analyzed with the use of mathematical models to determine adsorption isotherm and kinetics. The novelty of the work lies in the selected combination of the lattice as well as in the adopted synthesis approach. Compared to available literature, a facile approach has been adopted in the present work where no toxic precursors and extreme reaction conditions were employed to synthesize the material. The synthesized material exhibited high selectivity toward anionic azo dyes and showed fast uptake of CR dye (within 15 min) with high removal efficiency of 96% at neutral pH conditions in comparison to the similar lattices reported in literature.[31]

Experimental section

Chemicals

Copper acetate monohydrate (≥99.0%), EG [$(CH_2OH)_2$] (≥99%), and Zinc acetate dihydrate (99.5%) were purchased from Merck (Rahway, NJ).

Preparation of EG modified-Zn/Cu HDS

To 100 mL solution of water and EG (1:1), 1.2 g of Zinc acetate dihydrate and 0.42 g of copper acetate monohydrate were added and stirred. The reaction mixture was ultrasonicated for 30 min and resultant solution was kept at 120 °C

Table 1. Analyses of adsorption efficiencies of some sorbents for removal of CR dye.

Adsorbents	Adsorption capacity, q_m (mg g^{-1})	Percentage removal (%)	pH	References
Mg-Al-LDH	111.11	93.4	9	[26]
Dialdehyde microfibrillated cellulose/chitosan film	152.5	99.9	–	[27]
Shrimp shell powder	288.20	96	<10.6	[14]
Kenaf-based activated carbon	14.749	95	6	[28]
Carboxymethyl chitosan/phytic acid hydrogel	8.49	89.7	7	[29]
Xanthan gum-g-Polyacrylamide/SiO_2 nanocomposite	209.205	96.37	4	[30]

temperature in an oven to reduce it to 10 mL volume. The reaction mixture was washed with 1:1 solution of ethanol and milliQ water and then dried in oven at 60 °C to obtain the blue-colored product.

Characterization

The diffraction pattern of the synthesized lattice was recorded using an advanced Bruker D8 diffractometer over the range of $2\theta = 5$–$70°$, employing Cu Kα radiation ($\lambda = 1.5418$ Å) acquired *via* a gobel mirror with 1.0 s/step scan rate and 0.02° step size at 298 K. The sample was analyzed using a PerkinElmer Fourier-transform infrared spectrometer version 10.5.3 (Waltham, Massachusetts, US). Thermogravimetric analysis was conducted using the PerkinElmer TGA in the 50–900 °C temperature range under flowing nitrogen and at a uniform heating rate of 10 °C min^{-1}. Brunauer–Emmett–Teller (BET) surface area was recorded at 77 K using a Quanta chrome Nova-1000 instrument by N_2 adsorption-desorption isotherm. FE-SEM with EDX analysis of Zn/Cu HDS was performed using a Zeiss GeminiSEM. PHI 5000 VersaProbe III was used to record XPS measurements with Al Kα for photoelectron excitation. Samples were characterized by UV–Vis spectra *via* a Cary Series UV–Vis Spectrophotometer (Cary 300 UV–Vis).

Details of adsorption experiments

An adsorption study using EG-modified HDS was conducted in triplicate for CR dye uptake. To ascertain the impact of various factors on CR dye removal, batch studies were performed under dark conditions. Separation of residual solution from adsorbent was achieved by centrifuging the mixture at 3,000 rpm. To record the dye solution concentration, a double-beam UV–Visible spectrophotometer was employed at the λ_{max} of 497 nm (for CR dye). The removal percentage of dye is estimated using the formula given in Eq. (1) whereas the quantity of CR dye adsorbed at equilibrium and time, *t*, i.e., q_e and q_t in mg g^{-1} are calculated according to Eqs. (2) and (3), respectively.

$$\% \text{ Removal} = \frac{C_o - C_e}{C_o} \times 100 \quad (1)$$

$$q_e = \frac{C_o - C_e}{W} \times V \quad (2)$$

$$q_t = \frac{C_o - C_t}{W} \times V \quad (3)$$

The dye concentration (mg L^{-1}) at initial and equilibrium stages of adsorption is represented by C_o and C_e, respectively, concentration at any time *t* is denoted as C_t, *W* is the adsorbent amount in mg and V denotes the solution of dye solution (in mL).

Results and discussion

Structural, morphological and thermal behavior

The formation of EG-modified Zn/Cu HDS was first confirmed using PXRD pattern exhibiting strong reflection with high intensity at a lower 2θ value (10.7°) and weak reflections at a higher value of 2θ around 21°, 32°, 36°, 44° and 57° suggesting the formation of layered structure of Zn/Cu HDS (Fig. 1(A)). The observed reflections can be indexed in rhombohedral symmetry having $a = 3.15$ and $c = 25.77$ Å. No diffraction peaks corresponding to zinc oxide or copper oxide can be observed suggesting the formation of lattice in pure phase. The existing literature reports that the appearance of first intense reflection at (003) is attributed to the summation of interlamellar distance and thickness of the brucite-like sheet (4.8 Å).[32] The observed gallery height of synthesized EG-modified Zn/Cu hydroxide is 3.42 Å which is close to the thickness of the EG layer (4.2 Å).[33] Therefore, suggesting the monolayer arrangement EG in the interlamellar region of Zn/Cu HDS.[34]

The existence of EG in the interlayer region was further analyzed using FTIR spectroscopy (Fig. 1(B)). The appearance of bands at 2,883 and 2,834 cm^{-1} are due to the antisymmetric and symmetric vibrations of the $-CH_2$ group of the EG molecule, respectively. The shift in bands toward lower wavenumber with significant decrease in intensity of C–H stretching vibrations in comparison to pure EG suggest the intercalation of EG in the interlamellar region. Additionally, the broad and sharp bands at 3,434 and 1,574 cm^{-1} signified the stretching and bending vibrations of hydroxide groups indicating hydrogen bonding between them. A series of low-intensity absorption bands at 1,412, 1,071 and 908 cm^{-1} are ascribed to CH_2 wagging, C–O–H bending, and CH_2 rocking vibrations of EG anions, respectively.[35,36] The original bands for CH_2 rocking vibration of EG molecules are observed at 882 and 864 cm^{-1} but here blue shift can be observed, i.e., shift to a higher wavenumber (908 cm^{-1}) with a single broad band which might be due to the increased bond order of C–C bond in EG.[36] The presence of metal-oxygen linkages can be confirmed from the bands present between 500 and 800 cm^{-1}.

The thermal stability of EG-modified Zn/Cu HDS was assessed from TGA thermogram presented in Figure 1(C). Based on the thermogravimetric trace of the sample, the loss of weight was seen in three distinct stages, with an overall weight loss of 37%. The mass loss below 114 °C could be accredited to the removal of surface-adsorbed molecular water. The second weight-loss stage ranging from 114 to 237 °C accounted for the elimination of physically adsorbed water and interlayer EG molecules. Weight loss between 237 and 445 °C suggested EG combustion, along with dehydroxylation of the lattice or collapse of a layered structure with the release of other by-products.[37,38]

The formation of layered structure and presence of metal ions was confirmed using SEM/EDX analysis (Fig. 2(A,B)). The SEM micrograph shows sheet-like morphology and EDX analysis confirmed the presence of zinc, oxygen and

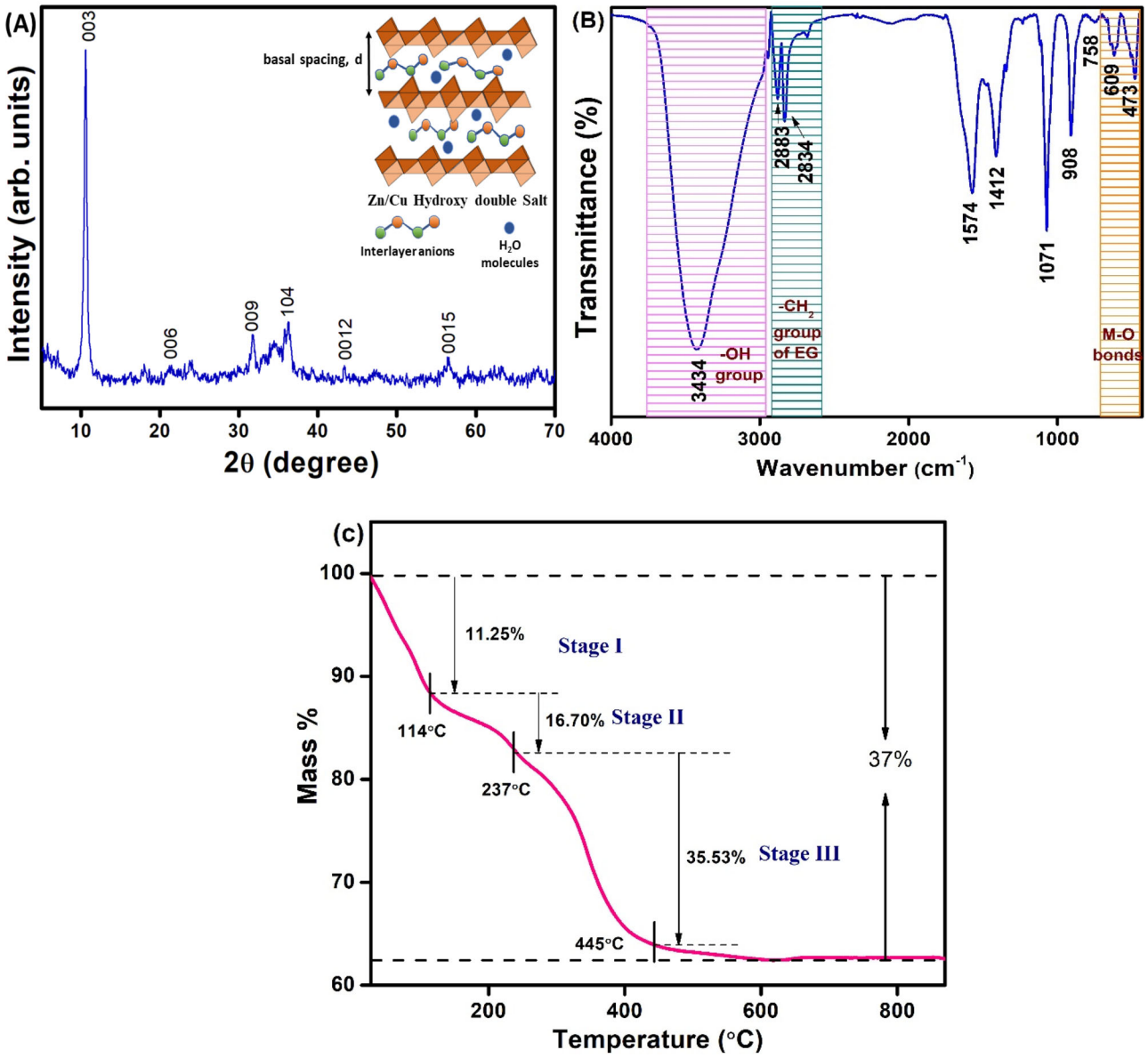


Figure 1. (A) PXRD pattern, (B) FTIR spectrum and (C) Thermogravimetric trace of EG-modified Zn/Cu HDS.

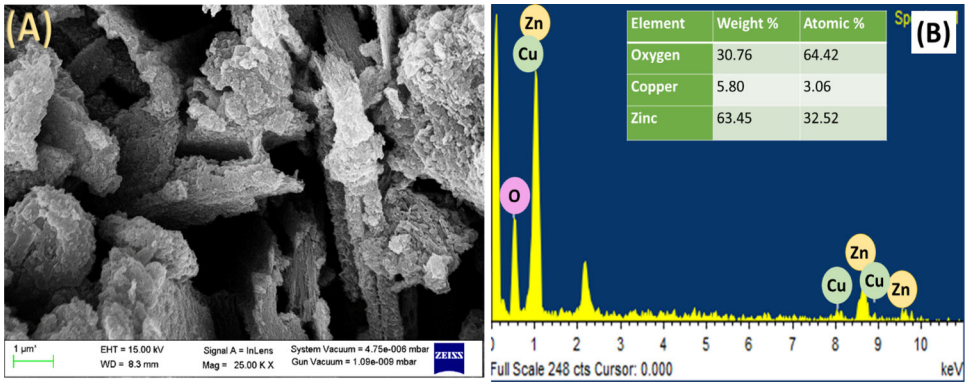


Figure 2. (A) SEM image and (B) EDX spectra of EG-modified Zn/Cu HDS.

copper elements and the percentage of each element is given in EDX spectrum[25]. The elemental mapping images are shown in Figure 3(A–E) which revealed the uniform distribution of elements in the lattice.

Further, the oxidation states and distribution of the elements were estimated by analyzing XPS data. The survey spectrum of XPS shown in Figure 4(A) indicated the presence of zinc, copper, oxygen and carbon elements. The

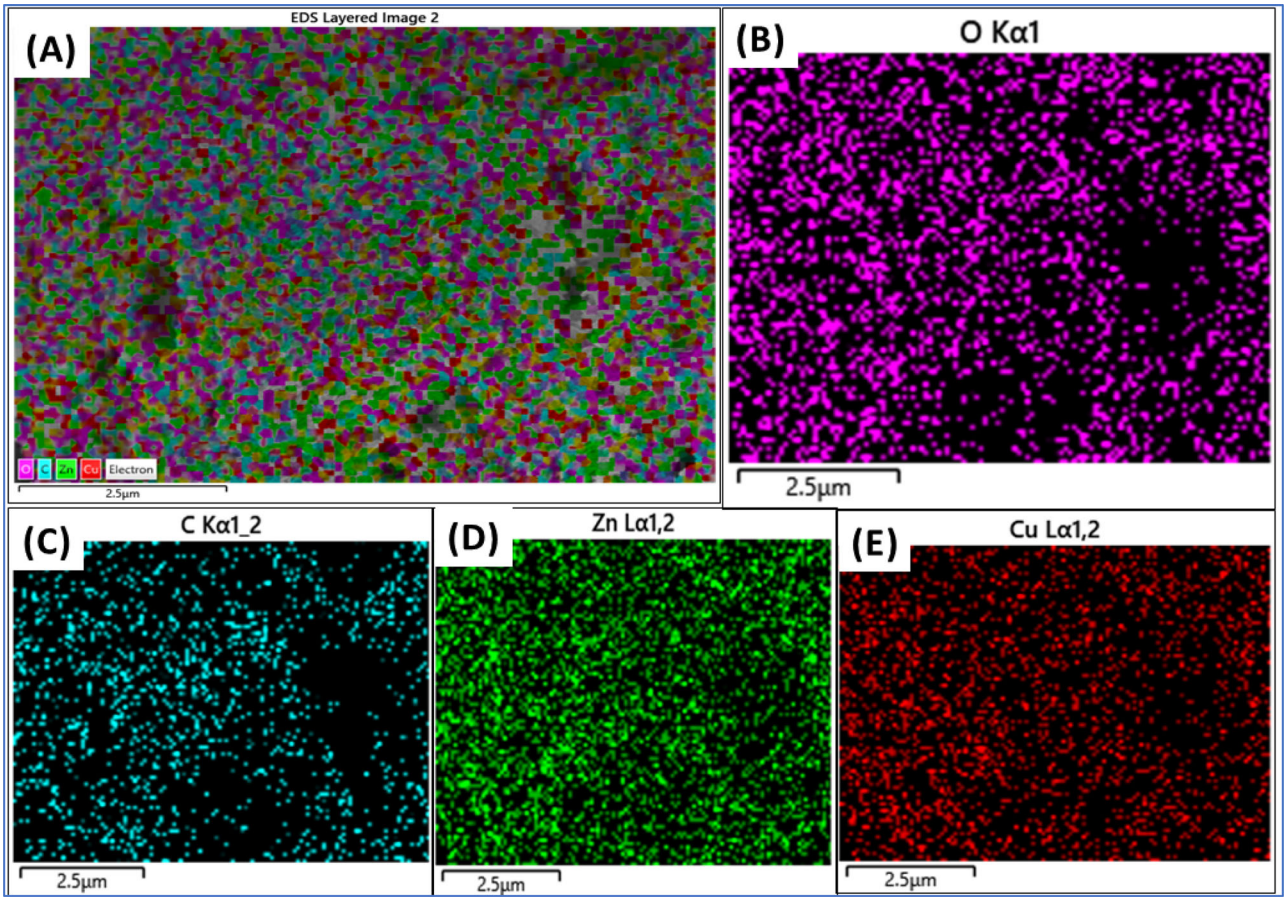


Figure 3. (A – E) corresponding elemental mapping of O, C, Zn and Cu, respectively, of EG-modified Zn/Cu HDS.

deconvoluted spectrum of C *1s* spectrum (Fig. 4(B)) shows two peaks centered at 284.96 and 288.30 eV ascribed to the C–C and C–O bonds of EG, respectively.[39] The deconvoluted O *1s* spectrum (Fig. 4(C)) exhibits two peaks at 529.94 and 531.44 eV, corresponding to the C–OH group of EG and metal-oxygen linkage (M–O), respectively.[40] The presence of two prominent peaks at 1,021.40 and 1,044.50 eV in core spectra of Zn *2p* (Fig. 4(D)) confirmed the presence of Zinc in the lattice.[41] The core level XPS spectrum of Cu *2p* (Fig. 4(E)) shows peaks with binding energies of 932.64 and 952.42 eV, confirming the presence of copper in +2 oxidation state in the lattice.[42]

BET surface area analysis

The porosity of EG-modified Zn/Cu HDS sample was examined using an N_2 adsorption-desorption isotherm. According to BET calculations, the adsorbent has a specific surface area of 32.295 m^2 g^{-1}, whereas the average pore size diameter and pore volume evaluated by Barrete–Joynere–Halenda (BJH) method are 2.218 nm and 0.070 cm^3g^{-1}, respectively (Fig. 5(A)). The isotherm depicts a type-IV isotherm with H_3 hysteresis loop, which is suggestive of the material's mesoporous nature.[43] The H_3 hysteresis is a feature of materials with slit-shaped pores.[44] BJH pore size distribution indicated that the lattice was composed of pores ranging from 2.13 to 3.74 nm, as shown in Figure 5(B). According to these findings, the porous nature of the synthesized material contributes greatly to its excellent sorption properties. With large surface areas and small pore sizes, mesoporous channels offer a sufficient surface to adsorb molecules of large size, such as organic dyes. As a result of the mesoporosity of the material, its potential as a sorbent for CR dye uptake from solution phase has been examined.

Adsorption properties

In order to assess the efficiency of synthesized lattice for the removal of various dyes from aqueous solution, adsorption experiment was performed. For this study, two cationic (crystal violet [CV] and methylene blue [MB]) and three anionic dyes (CR, MO and eriochrome black T [EBT]) were used. The removal of different dyes was performed with 20 mg L^{-1} concentration of dye with a fixed amount of adsorbent. Figure 6(A) shows the plot for removal of different dyes and it can be observed that it follows the trend: CR (96%) > EBT (77%) > MO (70%) > CV (13%) > MB (2%). From the order, it can be deduced that the synthesized material exhibits higher affinity for the anionic dyes which can be attributed to the fact that Zn/Cu HDS possess excess positive charge on the layers and there occurs an electrostatic interaction between the positively charged layers and anionic dye molecules. In addition, selective adsorption of CR dye was also performed in presence of cationic dye, i.e., MB under similar reaction conditions. It was observed that

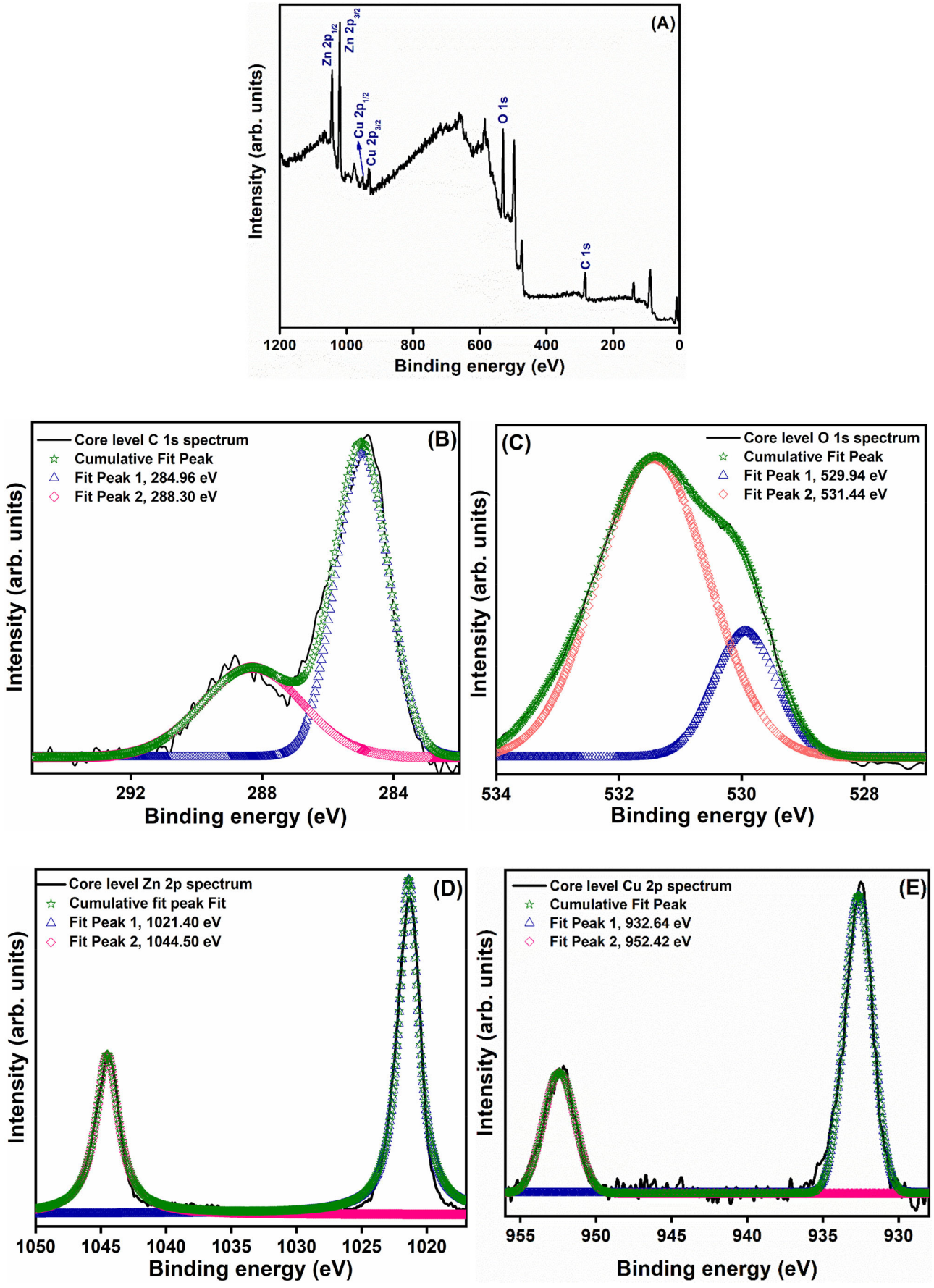


Figure 4. (A) Survey XPS spectrum of EG-modified Zn/Cu HDS, (B), (C), (D) and (E) show the XPS spectrum of C *1s*, O *1s*, Zn *2p* and Cu *2p*, respectively.

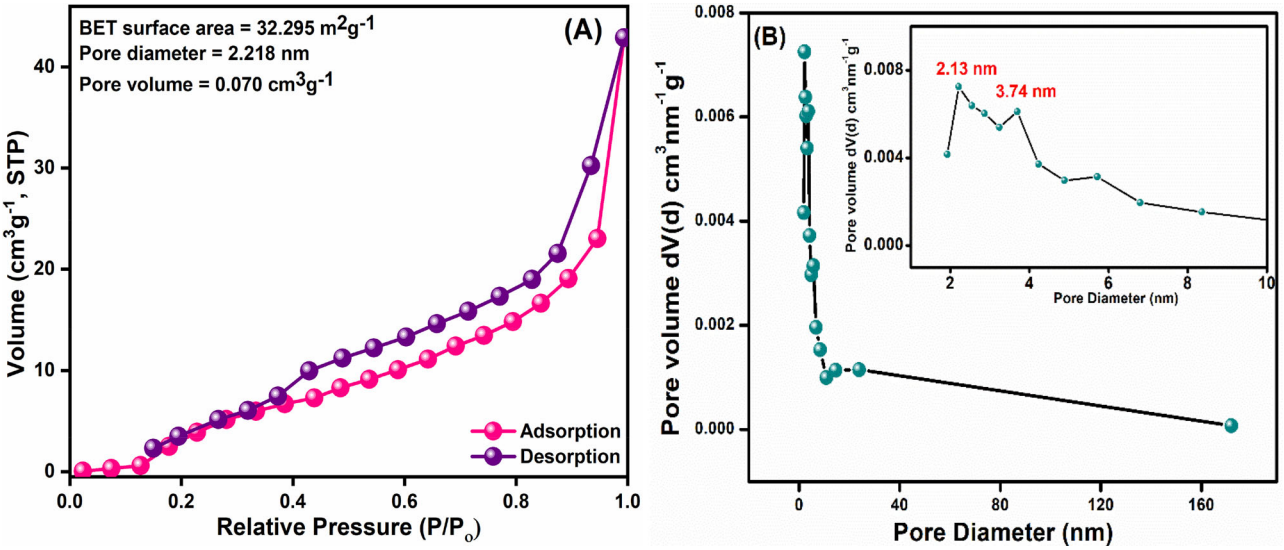


Figure 5. (A) N_2 Adsorption-desorption isotherm and (B) Plot of BJH pore size distribution of EG-modified Zn/Cu HDS.

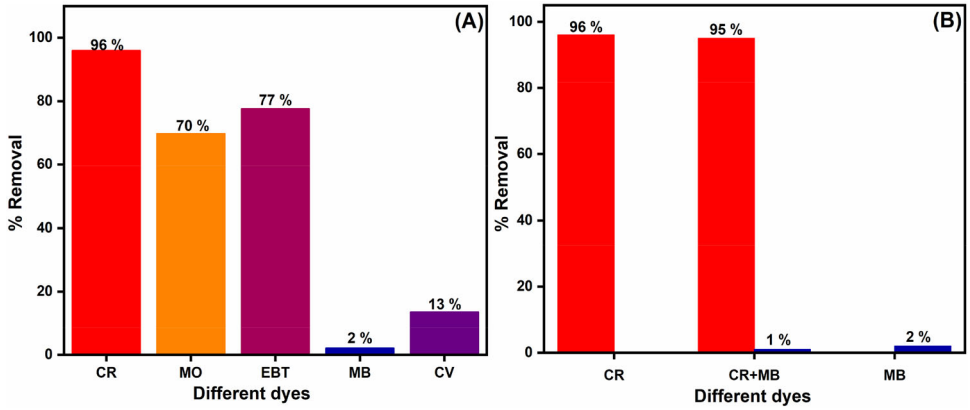


Figure 6. Plot for (A) % removal of different dyes and (B) selective adsorption of CR dye.

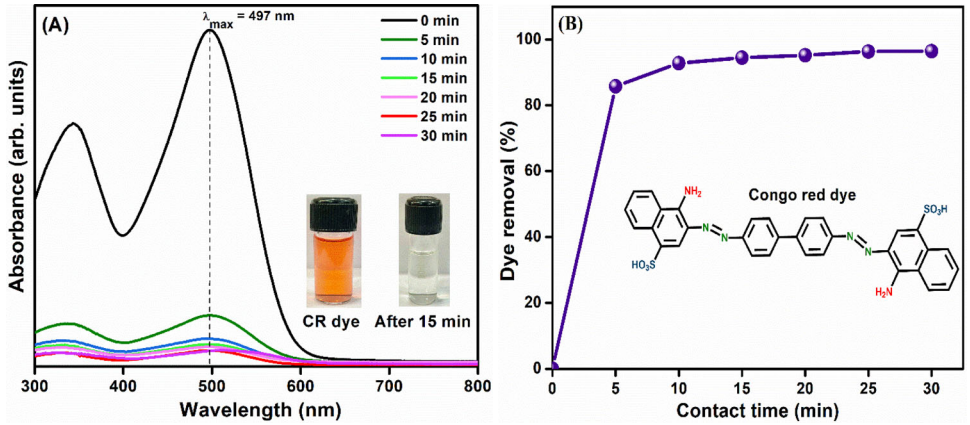


Figure 7. Plot of (A) absorbance vs wavelength and (B) influence of contact time on dye uptake efficacy of EG-Zn/Cu HDS.

the synthesized adsorbent exhibits high selectivity toward the anionic dye with a removal efficiency of 96%. The results obtained are shown in Figure 6(B).

Influence of contact time

Treatment of wastewater by the adsorption route relies significantly on the contact time between adsorbents and pollutants. To investigate the impact of contact time, 0.05 g of EG-modified Zn/Cu HDS product was transferred to 100 mL aqueous solution of CR dye with 20 mg L^{-1} concentration. The plot of absorbance *vs.* contact time is depicted in Figure 7(A). CR dye uptake was shown to have greatly increased with time and subsequently reached equilibrium afterward. A significant increase in dye adsorption occurred during the first 5 min of reaction time as shown in

Figure 7(B), reaching >90%, following which equilibrium was achieved (after 10 min). More than 96% of the CR is removed with 0.05 g of adsorbent, and it remains almost constant as contact time increases, suggesting that equilibrium has been reached. This was mainly because of the large area of surface area available for adsorption during the initial 5 min of contact time. Subsequently, the number of vacant sites decreases as the contact time increases because of repulsive interactions between solid and bulk phases.

Influence of dye concentration

The significance of concentration of dye was studied for aqueous solutions of 20–100 mg L^{-1} concentrations of dye with 0.05 g of adsorbent for 30 min. From Figure 8(A), one can conclude that adsorption capacity increases with an increase in dye concentration. On increasing CR dye concentration (20–100 mg L^{-1}), a significant increase in adsorption efficiency was obtained (39.23–169.18 mg g^{-1}). A higher molecular concentration increases the likelihood of molecular collisions with adsorbent surfaces, which in turn contributes to an increase in mass transfer from solution to solid. The increasing number of collisions results in increased removal of molecules from the solution phase. In summary, the initial high concentration of dye presents a greater concentration slope or a stronger driving force among the liquid phase (dye solution) and adsorbent phase, consequently, the adsorption rate increases.

Influence of amount of adsorbent

To inspect the impact of different amounts of adsorbent on the dye removal, trials were performed with different dosages of adsorbent (0.05 g, 0.1 g and 0.2 g) while keeping other parameters constant (initial dye concentration = 20 mg L^{-1} and reaction time = 30 min). Figure 8(B) shows a trend that adsorption efficiency enhanced on increasing amount of adsorbent, confirming that the greater the number of active sites for sorption greater will be the removal.

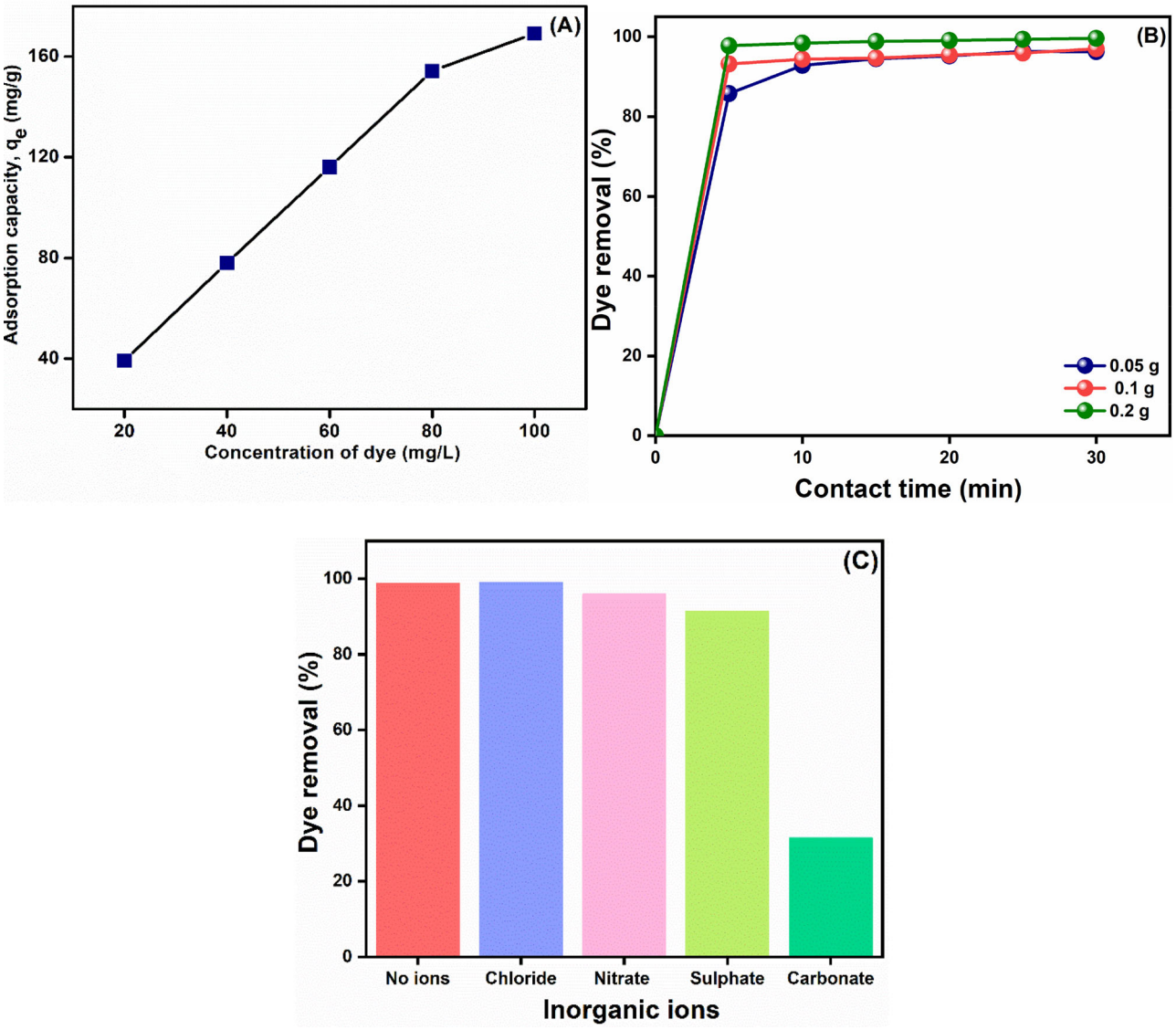


Figure 8. Influence of (A) concentration variation on adsorption capacity, (B) adsorbent dosage and (C) interfering anions on % dye adsorption.

Influence of interfering anions

Textile dyeing processes generate a large amount of toxic effluent, which generally contains various inorganic salts, such as sodium chloride, sodium carbonate, etc.[45] To estimate the dye removal efficacy of EG-modified Zn/Cu HDS as an adsorbent, the adsorption procedure needs to be analyzed to determine the impact of these anions. During this study, chloride (Cl^-), carbonate (CO_3^{2-}), nitrate (NO_3^-) and sulfate (SO_4^{2-}) anions were investigated and the adsorption experiment was carried out by taking 0.05 M concentration of each anion.

According to Figure 8(C), the presence of Na_2SO_4 and Na_2CO_3 to dye solutions resulted in a decrease in dye uptake, as opposed to the presence of NaCl, $NaNO_3$, or none of the electrolytes. It has been reported that divalent species have a high affinity for the layered materials and thus compete with dye molecules for adsorption sites on material's surface resulting in decreased adsorption of anionic dye molecules. Since, Cl^- and NO_3^- are monovalent ions they did not influence the adsorption efficiency significantly. However, a substantial decrease in removal efficiency of the material is demonstrated by divalent anions. Carbonate ions are negatively charged, similar to anionic dyes, can effectively interact with the positively charged sites on the layered material, thereby occupying these sites and reducing the availability for the dye molecules. A decrease in the adsorption efficiency of the material occurs as a result of the competition between carbonate ions and anionic dyes for adsorption sites. As opposed to carbonate ions, sulfate ions are relatively larger in size, which may cause less effective blocking of adsorption sites on layers. Therefore, sulfate ions do not significantly alter materials' adsorption efficiency. Owing to the competitive uptake of dye molecules and inorganic salts on surface of adsorbent, dye solutions containing inorganic salts have a decreased adsorption efficiency. Consequently, it can also be stated that the presence of divalent ions strongly influences the adsorptive removal of CR dye as compared to monovalent ions, which is in agreement with the fact that HDS has a high affinity for anions having high charge density.[46]

Influence of pH on dye adsorption

When it comes to water treatment applications, dye solution pH is a key factor in regulating the efficacy of adsorbents. To analyze the influence of pH, adsorption studies were conducted at different pH (pH 3–11) with 20 mg L^{-1} concentrations of dye solutions and 0.05 g of adsorbent dosage for 30 min. Figure 9(A) depicts the dye removal percentage at different pH values, it was observed to decrease from 99.54% to 21.62% when the pH was varied from 3 to 11 and maximum removal was recorded at pH 3 and 7. In general, the adsorption of anionic dyes like CR decreases as pH increases. Since excess hydroxide ions contest with dye molecules for adsorption sites at high pH levels, dye adsorption capacity is found to reduce.

Point of zero charge (PZC) of the sorbent can explain the decline in adsorption efficiency with increased pH. PZC is the pH at which a material's surface becomes neutral. When pH of the material is below pH_{PZC}, the material's surface exhibits a positive charge, while above pH_{PZC}, the material is negatively charged. In this study, the PZC was found to be 7.6 (Fig. 9(B)) indicating that the synthesized HDS exhibits a positively charged surface at pH $<$ 7.6, resulting in enhanced dye adsorption due to electrostatic interactions. As the pH of solution rises above 7.6, HDS hydroxide ions contest with anionic dye molecules causing a decline in the dye removal efficacy.

Proposed adsorption mechanism

HDSs are porous materials having numerous adsorption sites. The mechanisms involved in the adsorption of pollutants using layered materials generally include physical adsorption, ion exchange reactions, electrostatic interactions, pi–pi interactions and chemical bonding.[47] The zeta potential measurements of the adsorbent can be used to explain the mechanism of adsorption. In both pre- and post-adsorption analyses, the material showed zeta potential value of 9.79 mV and 4.95 mV, respectively. It is suggested that the surface of the adsorbent possesses a positive charge due to its positive zeta potential value. Conversely, CR has a high solubility in water and exists in anionic form under the aqueous phase. Thus, a decrease in value of zeta potential upon adsorption of CR can be attributed to the electrostatic

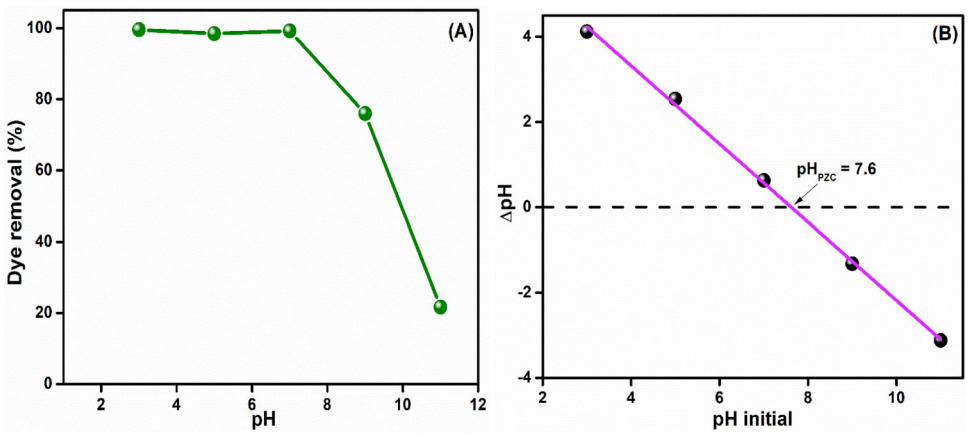


Figure 9. (A) pH effect on percentage dye removal and (B) plot of pH_{PZC}.

interactions occurring between the positive layers of adsorbent and negative dye molecules at neutral or low pH environments. Therefore, it can be stated that electrostatic interaction (between metal ions and dye molecules), as well as hydrogen bonding (between EG and dye molecules), are responsible for the high removal efficiency of synthesized adsorbent toward CR dye. The proposed mechanism of adsorption is depicted in Figure 10.

Adsorption isotherms

This work used three isotherm models to analyze the equilibrium adsorption experimental data for removing CR dye. The first model is Langmuir isotherm implying that sorption occurs on uniform, homogeneous surfaces with identical sites on an adsorbent and it is often used to describe monolayer adsorptions.[48] Following are the non-linearized (4) and linearized (5) forms of the model, respectively.

$$q_e = \frac{q_m k_L\ C_e}{(1 + k_L\ C_e)} \tag{4}$$

$$\frac{1}{q_e} = \frac{1}{q_m k_L\ C_e} + \frac{1}{q_m} \tag{5}$$

q_e and C_e are uptake capacity in mg g^{-1} and dye concentration at equilibrium in mg L^{-1}, respectively. Langmuir constants are represented by k_L and q_m defined as adsorption energy (L/mg) and monolayer or maximum adsorption capacity (mg g^{-1}), respectively. In accordance with Eq. (5), when the Langmuir equation holds, the graph of $1/q_e$ *vs.* $1/C_e$ should possess an intercept of $1/q_m$ and a slope of $1/q_m k_L$. R_L is the dimensionless factor that allows us to quantify Langmuir isotherm and can be defined as:

$$R_L = \frac{1}{1 + k_L C_0} \tag{6}$$

where initial stage concentration of dye (mg L^{-1}) is presented by C_0 and R_L value indicates the characteristics of the adsorption, i.e., favorable ($0 < R_L < 1$), unfavorable ($R_L > 1$), linear ($R_L = 1$) and irreversible ($R_L = 0$).

Another model is Freundlich isotherm which can be very useful for understanding the sorption mechanism for heterogeneous surfaces. The non-linearized and linearized forms of the model are represented in Eqs. (7) and (8), respectively.

$$q_e = k_F C_e^{\,1/n} \tag{7}$$

$$\ln q_e = \ln k_F + \frac{1}{n} \ln C_e \tag{8}$$

$1/n$ and K_F are constants expressing intensity and capacity of adsorption, respectively. The slope and intercept of a graph of *ln* q_e *vs. ln* C_e allow us to compute $1/n$ and k_F values. An adsorption process is characterized as favorable if the $1/n$ ratio is higher than 1, indicating favorable conditions for adsorption.

The indirect adsorbate–adsorbent interactions are taken into consideration in the third model, i.e., Temkin isotherm model. The increase in surface coverage is assumed to cause a linear decrease instead of a logarithmic in the heat of adsorption.[48] The nonlinearized-linearized form of Temkin adsorption model is presented in Eqs. (9) and (10).

$$q_e = B \ln (A_T C_e) \tag{9}$$

$$q_e = B \ln A_T + B \ln C_e \tag{10}$$

Temkin constants A_T and B are defined as the binding constant at equilibrium and heat of sorption, respectively.

Figure 11(A–D) illustrates the results for the different adsorption isotherms utilized in the present analysis. The relevant parameters of these isotherms are also given in Table 2. There is also good agreement between the results obtained from linear and nonlinear isotherm models.

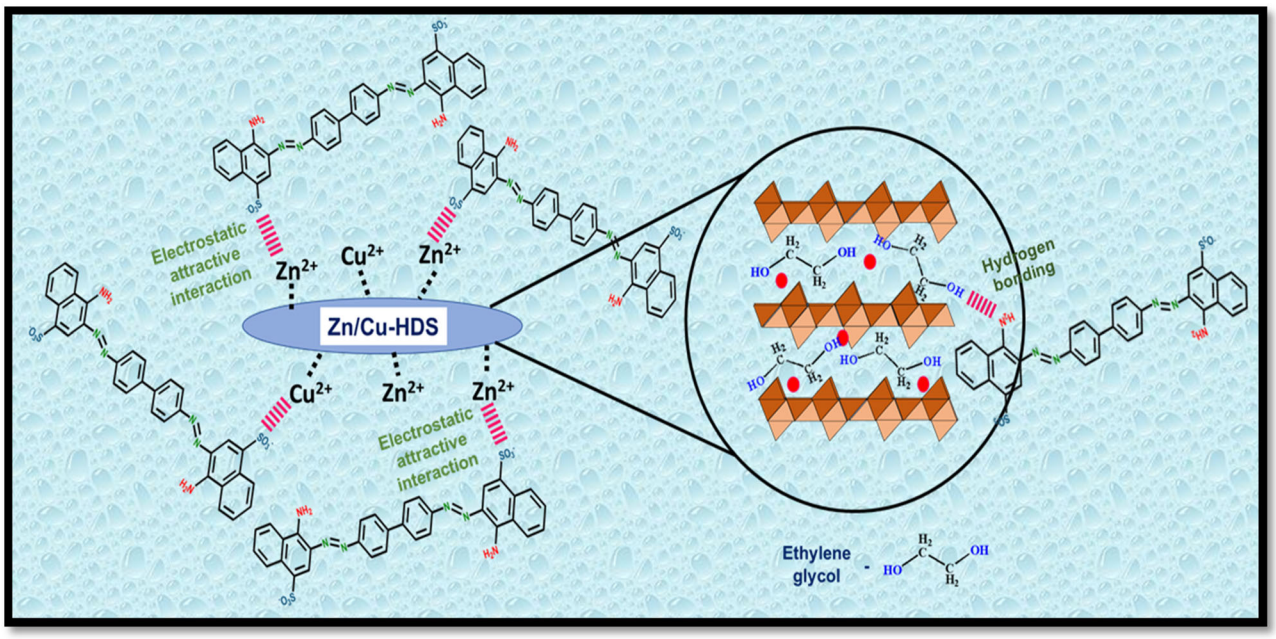


Figure 10. Plausible mechanism of adsorption of CR dye onto EG-Zn/Cu HDS.

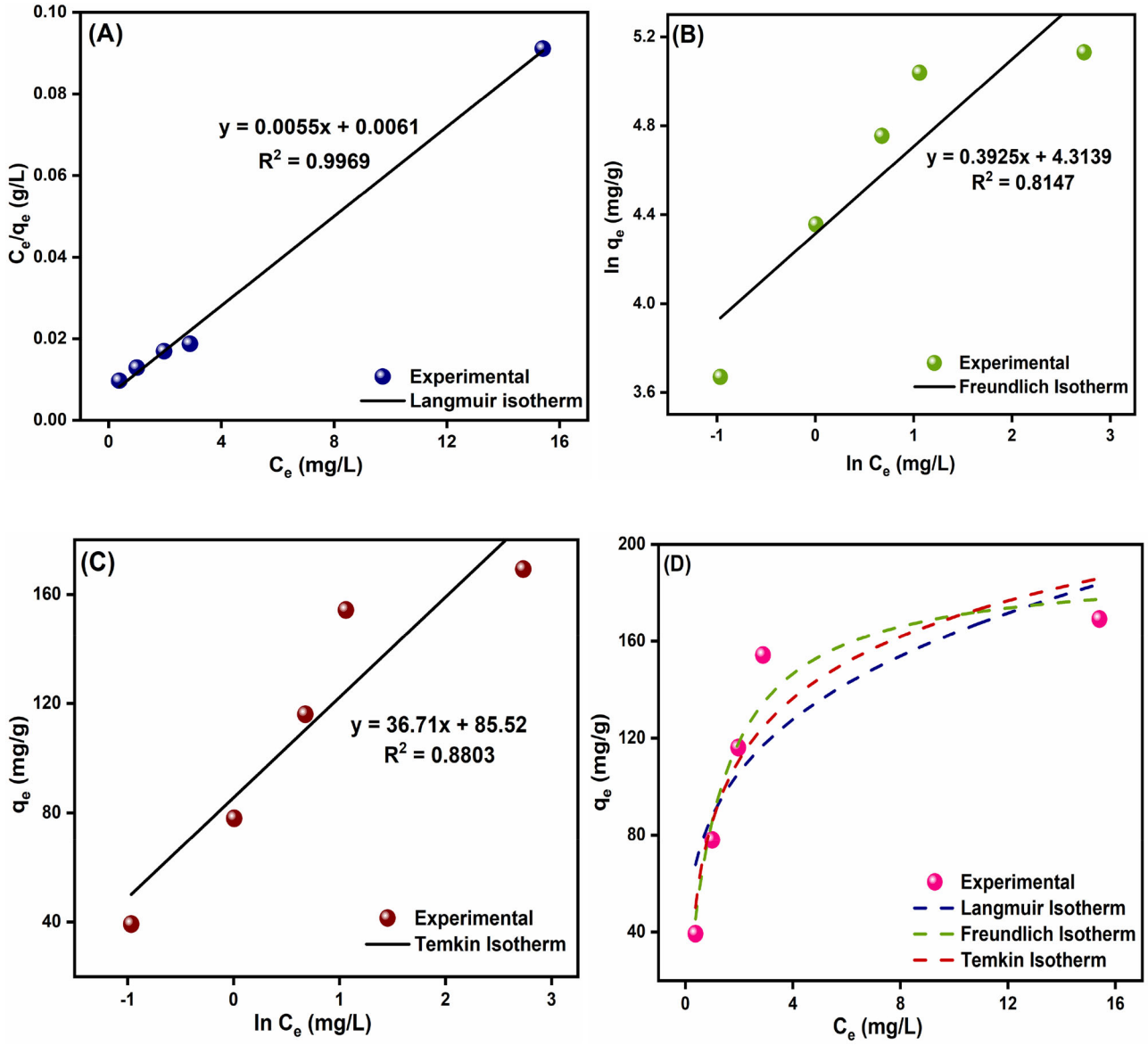


Figure 11. Isotherm linear plots for (A) Langmuir, (B) Freundlich, (C) Temkin model and (D) nonlinear plots of isotherm models for removal of CR dye using EG-Zn/Cu HDS.

Table 2. Adsorption isotherm constants for CR dye removal.

Equilibrium isotherms	Equations	Parameters		
Langmuir	$\frac{1}{q_e} = \frac{1}{q_m k_L}\frac{C_e}{C_e} + \frac{1}{q_m}$	q_m (mg g^{-1})	K_L (L/mg)	R^2
Linearized form	$q_e = \frac{q_m k_L C_e}{(1+k_L C_e)}$	181.81	0.9016	0.9969
Non-linearized form		191.55	0.8135	0.9504
Freundlich	$\ln q_e = \ln k_F + \frac{1}{n}\ln C_e$	K_F (mg g^{-1}) (L/mg)$^{1/n}$	*1/n*	R^2
Linearized form	$q_e = k_F C_e^{1/n}$	74.73	0.3925	0.8147
Non-linearized form		87.90	0.2692	0.7720
Temkin	$q_e = B\ln A_T + B\ln C_e$	B (J/mol)	A_T (L/g)	R^2
Linearized form	$q_e = B\ln(A_T C_e)$	36.71	10.26	0.8803
Non-linearized form		36.71	10.26	0.8803

By examining the values of R^2, it is evident that the calculated adsorption data agrees with the Langmuir isotherm, indicating monolayer adsorption of CR dye onto EG-Zn/Cu HDS adsorbent. In addition, the R_L was 0.90, indicating the favorable adsorption of CR dye.

Kinetic studies

For designing and modeling adsorption processes, kinetic parameters can be used to predict adsorption rates. To investigate the adsorption mechanism, the obtained data were then correlated with three different kinetic models.

Adsorption rates can be demonstrated in terms of adsorption capacity based on linear and nonlinear Lagergren's first-order rate equations represented as follows:

$$q_t = q_e(1 - e^{-k_1 * t}) \tag{11}$$

$$\ln(q_e - q_t) = \ln q_e - k_1 t \tag{12}$$

q_e and q_t (mg g^{-1}) are the uptake capacity at equilibrium and time t, respectively, while pseudo-first-order rate constant, k_1 (min^{-1}) can be computed from linear plot of ln (q_e – q_t) *vs.* *t*.

Ho and McKay proposed that a pseudo-second-order model could also be utilized to explain the kinetics of adsorption.[49] The linearized Eq. (13) for the kinetic model is given below:

$$\frac{t}{q_t} = \frac{1}{k q_e^2} + \frac{t}{q_e} \tag{13}$$

Which can be rearranged into its nonlinear form:

$$q_t = \frac{q_e^2 k_2 \ t}{q_e k_2 \ t + 1} \tag{14}$$

q_e and q_t (mg g^{-1}) represent adsorption capacity at equilibrium and at time t, respectively, whereas pseudo-second-order rate constant is presented by k_2 (g mg^{-1} min^{-1}) which can be computed from graph of t/q_t *vs.* *t*. Chemical adsorption is considered to be rate-limiting step in pseudo-second-order model that predicts the behavior of adsorption across the overall range.

Moreover, the adsorption mechanism was also examined using intra-particle diffusion model. The linearized form of the model presented by Weber and Morris is formulated in equation.[50]

$$q_t = k_{id} \ t^{1/2} + C \tag{15}$$

where the dye uptake capacity at time t and intra-particle diffusion rate constant are denoted by q_t (mg g^{-1}) and k_{id}.

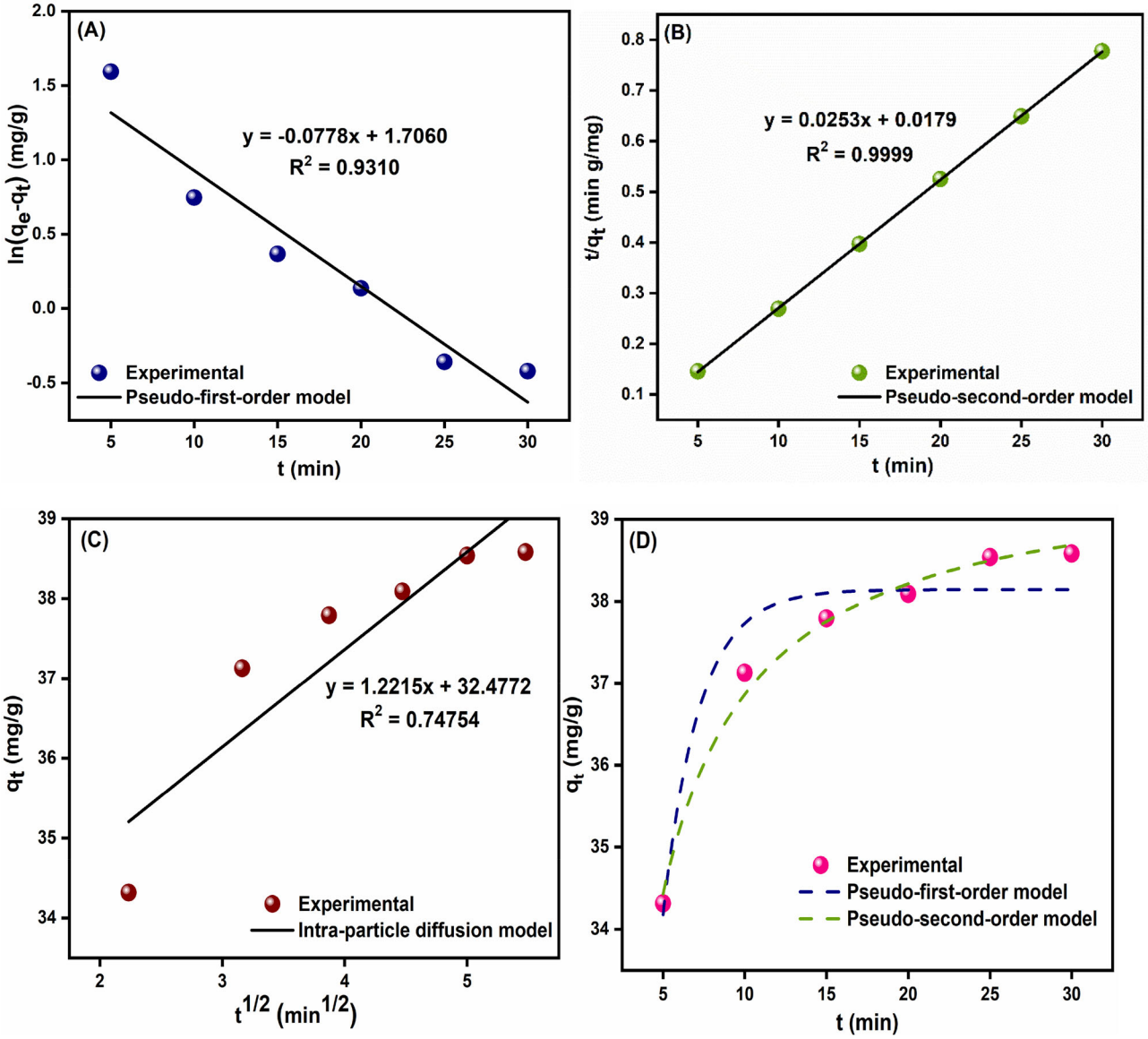


Figure 12. Linear plots of (A) Pseudo-first-order, (B) Pseudo-second-order, (C) Intra-particle diffusion and (D) nonlinear plots of kinetic models for removal of CR dye using EG-Zn/Cu HDS.

Table 3. Kinetic parametric values for CR dye uptake.

Kinetic models	Equations	Parametric values		
Pseudo-first order	$\ln(q_e - q_t) = \ln q_e - k_1 t$	K_1 (min^{-1})	q_e (mg g^{-1})	R^2
Linearized form	$q_t = q_e(1 - e^{-k_1 * t})$	0.0778	5.50	0.9310
Non-linearized form		0.0761	38.14	0.9359
Pseudo-second order	$\frac{t}{q_t} = \frac{1}{kq_e^2} + \frac{t}{q_e}$	K_2 (g mg^{-1} min^{-1})	q_e (mg g^{-1})	R^2
Linearized form	$q_t = \frac{q_e^2 k_2 q_e t}{q_e k_2 \ t+1}$	0.0331	39.52	0.9999
Non-linearized form		0.0357	39.66	0.9912
Intra-particle diffusion	$q_t = k_{id}\ t^{1/2} + C$	K_{id} (g/mg $min^{-0.5}$)	C	R^2
Linearized form		1.2215	32.47	0.7475

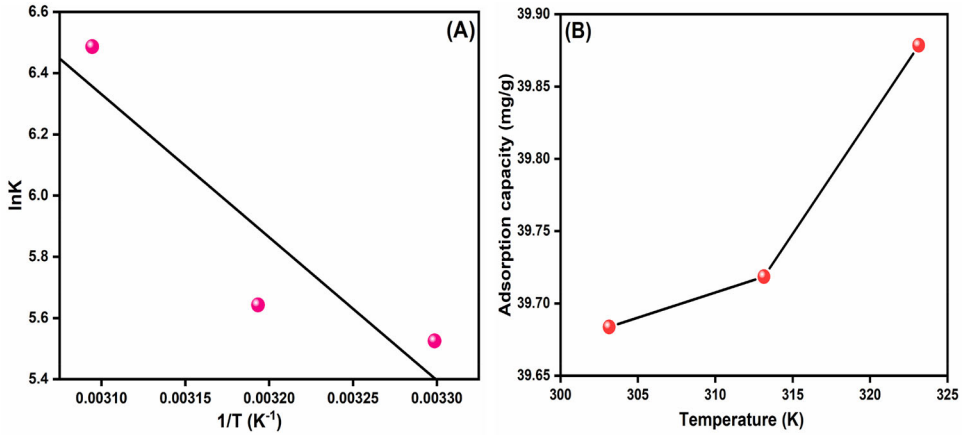


Figure 13. (A) Van't Hoff plot for ln *K vs.* 1/*T* and (A) plot for effect of temperature on adsorption capacity.

The results for kinetic models employed in this study are depicted in Figure 12(A–D). The kinetic variables for all the models are provided in Table 3. It was evident that pseudo-second-order model with 0.9999 R^2 value fits well with experimental data. In addition, the linearized and nonlinearized forms of the model also produce similar results. Moreover, the uptake capacity, q_e computed from pseudo-second-order model (39.52 mg g^{-1}) is in agreement with the observed value (39.23 mg g^{-1}). These results indicate CR adsorption follows a pseudo-second-order kinetic process, where chemical interactions enable CR to adsorb on the surface of synthesized HDS.

Effect of temperature

During adsorption, entropy and energy are important factors to consider when determining the spontaneity of the process. In order to obtain thermodynamic parameters for the adsorption system, the amount of CR adsorbed at equilibrium at different temperatures, i.e., 303.15, 313.15 and 323.15 K was examined. The following equations were used to estimate changes in Gibbs free energy, enthalpy and entropy of adsorption:

$$\ln K = -\frac{\Delta H^0}{RT} + \frac{\Delta S^0}{R} \tag{16}$$

$$K = \frac{q_e}{C_e} \tag{17}$$

$$\Delta G^0 = \Delta H^o - T\Delta S^0 \tag{18}$$

The adsorbates' distribution coefficient, gas constant, and absolute temperature are denoted by K, R (8.314 J/K/mol), and T, respectively. The values of ΔH° and ΔS° can be calculated from the linear plot of ln K *vs.* 1/T (Fig. 13(A)). The parameters calculated are given in Table 4. The positive value of change in enthalpy demonstrated that the adsorption process is endothermic and during adsorption, there is a higher degree of disorderness at the solid-solution interface. It was also observed that the adsorption capacity increased with an increase in temperature (Fig. 13(B)). The collision rate of dye and Zn/Cu HDS is increased at elevated temperatures because of the increase in kinetic energy of adsorbate molecules. In addition, the negative value of ΔG° confirms that the process is spontaneous and the increasing value of ΔG° with increasing temperature suggests that the rate of adsorption is higher at high temperatures.

Table 4. Thermodynamic parametric values for CR dye uptake.

Temperature (K)	q_e (mg g^{-1})	ΔG° (KJ/mol)	ΔH° (KJ/mol)	ΔS° (J/mol/K)
303.15	39.68	−13.61	38.83	173.02
313.15	39.72	−15.35	–	–
323.15	39.88	−17.07	–	–

Regeneration and stability of adsorbent

In order to investigate synthesized adsorbent's stability, regeneration ability and reusability for removing CR dye from aqueous phase a series of experiments were conducted under same reaction conditions for up to seven cycles. Following adsorption, the EG-Zn/Cu HDS adsorbent was centrifuged, washed with ethanol and dried, then reused for the subsequent cycles. The recovered adsorbent was found to be reusable for up to five cycles without significant loss

in its dye uptake efficiency. The plot for removal efficiency for consecutive five cycles is presented in Figure 14(A). Adsorption efficiency for CR dye uptake in the first cycle was 96.06%, and it decreased slightly in the subsequent cycles and reached 92.86% after 4 cycles whereas, the percentage removal decreased further in the 5th cycle and it was found to be 85. 98%. In light of these results, EG-Zn/Cu HDS serves as a potential candidate for dye-loaded wastewater remediation due to its stability and reusability for up to 5 cycles. The stability of the adsorbent after adsorption was confirmed from the PXRD results (before and after adsorption). It can be observed from Figure 14(B) that the XRD patterns of adsorbent before and after adsorption are similar with a slight change in the intensity of peaks and no shift in the peak position, suggesting that the structure of material is retained even after the adsorption process. The results confirmed the stability of the synthesized adsorbent.

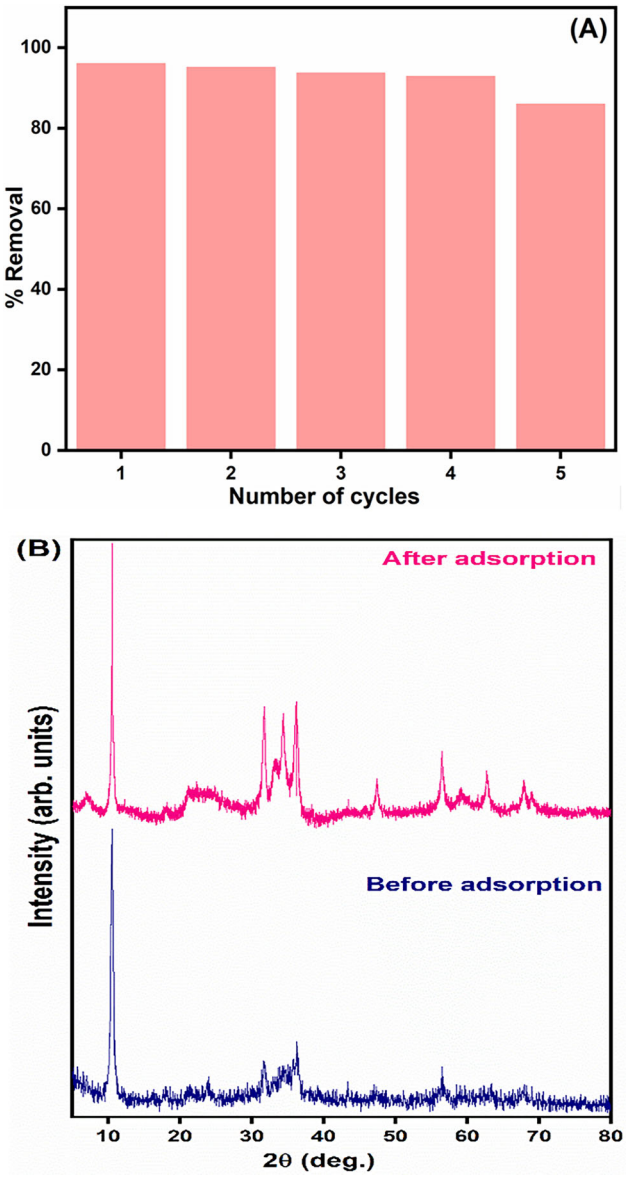


Figure 14. (A) Reusability of EG-Zn/Cu HDS for Congo red removal and (B) PXRD patterns of adsorbent before and after adsorption.

Conclusion

This study reports the synthesis of EG-modified HDS containing zinc and copper metal ions using a facile one-pot synthesis approach. The synthesized lattice was characterized using various characterization techniques. Additionally, BET analysis was used to analyze the textural properties of synthesized lattice and determine its surface area of 32.295 $m^2 g^{-1}$. The synthesized EG-modified HDS act as an effective adsorbent for the sequestration of CR dye under broad pH range of 3–7 with high removal efficiency of more than 97%. The experimental data best fitted Langmuir isotherm and pseudo-second-order kinetic model suggesting monolayer adsorption and chemisorption to be the rate-limiting step, respectively. Furthermore, EG-modified HDS displayed a maximum monolayer sorption capacity of 181.81 mg g^{-1}. These findings suggest that the EG-Zn/Cu HDS adsorbent could serve as a useful alternative for removing CR dye from aqueous solutions. In addition, the synthesized adsorbent was successfully used for five recycled runs with high removal efficiency. According to the findings of this study, the EG-Zn/Cu HDS adsorbent could serve as a useful alternative for removing CR dye from the aqueous phase owing to its environmental friendliness and low cost. Consequently, this work opens new opportunities for toxic anionic azo dyes uptake using functionalized Zn/Cu HDS.

Acknowledgments

The authors acknowledge their deep gratitude to Delhi Technological University for its support in conducting the study. The authors thank UGC-BSR startup grant (30-416/2018(BSR)) for funding this project. The author BP is also grateful for the financial support provided by CSIR, India (SRF). Special thanks to Material Analysis and Research Centre, BIT, Bangalore, India for BET measurements.

Disclosure statement

The authors declare no conflict of interest.

Funding

This work was supported by Delhi Technological University, UGC-BSR startup grant (30-416/2018(BSR)) CSIR, India (SRF) and Material Analysis and Research Centre, BIT, Bangalore, India.

ORCID

Bhamini Pandey http://orcid.org/0000-0003-3146-0141
Poonam Singh http://orcid.org/0000-0001-6786-8419
Vinod Kumar http://orcid.org/0000-0002-8362-8009

Data availability statement

Data available upon request.

References

[1] Williams, G. R.; Hare, D. O. Towards Understanding, Control and Application of Layered Double Hydroxide Chemistry. *J. Mater. Chem.* **2006**, *16*, 3065–3074. DOI: 10.1039/b604895a.

[2] Centi, G.; Perathoner, S. Catalysis by Layered Materials: A Review. *Microporous Mesoporous Mater.* **2008**, *107*, 3–15. DOI: 10.1016/j.micromeso.2007.03.011.

[3] Kumar, P.; Abuhimd, H.; Wahyudi, W.; Li, M.; Ming, J.; Li, L.-J. Review—Two-Dimensional Layered Materials for Energy Storage Applications. *ECS J. Solid State Sci. Technol.* **2016**, *5*, Q3021–Q3025. DOI: 10.1149/2.0051611jss.

[4] Kuthati, Y.; Kankala, R. K.; Lee, C. H. Layered Double Hydroxide Nanoparticles for Biomedical Applications: Current Status and Recent Prospects. *Appl. Clay Sci.* **2015**, *112–113*, 100–116. DOI: 10.1016/j.clay.2015.04.018.

[5] Evans, D. G.; Duan, X. Preparation of Layered Double Hydroxides and Their Applications as Additives in Polymers, as Precursors to Magnetic Materials and in Biology and Medicine. *Chem. Commun. (Camb).* **2006**, *37*, 485–496. DOI: 10.1039/b510313b.

[6] Pathak, J.; Singh, P. Synthesis and Characterization of Ternary Layered Double Hydroxide Containing Zinc/Copper/Nickel and Its PANI Composite. *Polym. Compos.* **2022**, *43*, 7836–7844. DOI: 10.1002/pc.26895.

[7] Morioka, H.; Tagaya, H.; Karasu, M.; Kadokawa, J.; Chiba, K. Preparation of Hydroxy Double Salts Exchanged by Organic Compounds. *J. Mater. Res.* **1998**, *13*, 848–851. DOI: 10.1557/JMR.1998.0112.

[8] Kandare, E.; Hossenlopp, J. M. Hydroxy Double Salt Anion Exchange Kinetics : Effects of Precursor Structure and Anion. *J. Phys. Chem. B* **2005**, *109*, 8469–8475. DOI: 10.1021/jp0465433.

[9] Kaassis, A. Y. A.; Xu, S.; Guan, S.; Evans, D. G.; Wei, M.; Williams, G. R. Hydroxy Double Salts Loaded with Bioactive Ions : Synthesis, Intercalation Mechanisms, and Functional Performance. *J. Solid State Chem.* **2016**, *238*, 129–138. DOI: 10.1016/j.jssc.2016.03.019.

[10] Li, A.; Mu, X.; Li, T.; Wen, H.; Li, W.; Li, Y.; Wang, B. Formation of Porous Cu Hydroxy Double Salt Nanoflowers Derived from Metal–Organic Frameworks with Efficient Peroxidase-like Activity for Label-Free Detection of Glucose. *Nanoscale* **2018**, *10*, 11948–11954. DOI: 10.1039/c8nr02832j.

[11] Lu, J.; Xu, C.; Tian, Z.; Lu, J.; Lin, Y.; Shi, Z. Green Emission and Ag + Sensing of Hydroxy Double Salt Supported Gold Nanoclusters. *Nanoscale* **2016**, *8*, 5120–5125. DOI: 10.1039/c5nr07029e.

[12] Crini, G. Non-Conventional Low-Cost Adsorbents for Dye Removal: A Review. *Bioresour Technol* **2006**, *97*, 1061–1085. DOI: 10.1016/j.biortech.2005.05.001.

[13] Sharma, A.; Rasheed, S.; Mangla, D.; Choudhry, A.; Shukla, S.; Chaudhry, S. A. Cobalt Ferrite Incorporated Ocimum Sanctum Nanocomposite Matrix as an Interface for Adsorption of Organic Dyes: A Sustainable Alternative. *ChemistrySelect* **2023**, *8*, e202203709. DOI: 10.1002/slct.202203709.

[14] Zhou, Y.; Ge, L.; Fan, N.; Xia, M. Adsorption of Congo Red from Aqueous Solution onto Shrimp Shell Powder. *Adsorpt. Sci. Technol.* **2018**, *36*, 1310–1330. DOI: 10.1177/0263617418768945.

[15] Mohapatra, D. P.; Kirpalani, D. M. Advancement in Treatment of Wastewater: Fate of Emerging Contaminants. *Can. J. Chem. Eng.* **2019**, *97*, 2621–2631. DOI: 10.1002/cjce.23533.

[16] Sharma, A.; Mangla, D.; Chaudhry, S. A.; Shehnaz . Recent Advances in Magnetic Composites as Adsorbents for Wastewater Remediation. *J. Environ. Manage.* **2022**, *306*, 114483. DOI: 10.1016/j.jenvman.2022.114483.

[17] Pandey, B.; Singh, P.; Kumar, V. Photocatalytic-Sorption Processes for the Removal of Pollutants from Wastewater Using Polymer Metal Oxide Nanocomposites and Associated Environmental Risks. *Environ. Nanotechnol. Monit. Manag.* **2021**, *16*, 100596. DOI: 10.1016/j.enmm.2021.100596.

[18] Lu, Y.; Cai, Y.; Zhang, S.; Zhuang, L.; Hu, B.; Wang, S.; Chen, J.; Wang, X. Application of Biochar-Based Photocatalysts for Adsorption-(Photo)Degradation/Reduction of Environmental Contaminants: Mechanism, Challenges and Perspective. *Biochar* **2022**, *4*, 45. DOI: 10.1007/s42773-022-00173-y.

[19] Qiu, M.; Liu, L.; Ling, Q.; Cai, Y.; Yu, S.; Wang, S.; Fu, D.; Hu, B.; Wang, X. Biochar for the Removal of Contaminants from Soil and Water: A Review. *Biochar* **2022**, *4*, 1–25. DOI: 10.1007/s42773-022-00146-1.

[20] Zhang, Y.; Liu, H.; Gao, F.; Tan, X.; Cai, Y.; Hu, B.; Huang, Q.; Fang, M.; Wang, X. Application of MOFs and COFs for Photocatalysis in CO2 Reduction, H2 Generation, and Environmental Treatment. *EnergyChem* **2022**, *4*, 100078. DOI: 10.1016/j.enchem.2022.100078.

[21] Liu, X.; Verma, G.; Chen, Z.; Hu, B.; Huang, Q.; Yang, H.; Ma, S.; Wang, X. Metal-Organic Framework Nanocrystal-Derived Hollow Porous Materials: Synthetic Strategies and Emerging Applications. *Innovation (Camb).* **2022**, *3*, 100281. DOI: 10.1016/j.xinn.2022.100281.

[22] Sharma, A.; Mangla, D.; Choudhry, A.; Sajid, M.; Chaudhry, S. A. Facile Synthesis, Physico-Chemical Studies of Ocimum Sanctum Magnetic Nanocomposite and Its Adsorptive Application against Methylene Blue. *J. Mol. Liq.* **2022**, *362*, 119752. DOI: 10.1016/j.molliq.2022.119752.

[23] Choudhry, A.; Sharma, A.; Khan, T. A.; Chaudhry, S. A. Flax Seeds Based Magnetic Hybrid Nanocomposite: An Advance and Sustainable Material for Water Cleansing. *J. Water Process Eng.* **2021**, *42*, 102150. DOI: 10.1016/j.jwpe.2021.102150.

[24] Erol, K.; Köse, K.; Köse, D. A.; Sızır, Ü.; Tosun Satır, İ.; Uzun, L. Adsorption of Victoria Blue R (VBR) Dye on Magnetic Microparticles Containing Fe (II) – Co (II) Double Salt. *Desalin. Water Treat.* **2016**, *57*, 9307–9317. DOI: 10.1080/19443994.2015.1030708.

[25] Weeramonkhonlert, V.; Srikhaow, A.; Meejoo, S. Formation of Copper Hydroxy Double Salts Derived from Metal Oxides and Their Catalytic Activity in Degradation of Methyl Orange. *Ceram. Int.* **2019**, *45*, 993–1000. DOI: 10.1016/j.ceramint.2018.09.278.

[26] Lafi, R.; Charradi, K.; Amine, M.; Ben, A.; Amara, H.; Hafiane, A. Adsorption Study of Congo Red Dye from Aqueous Solution to Mg – Al – Layered Double Hydroxide. *Adv. Powder Technol.* **2016**, *27*, 232–237. DOI: 10.1016/j.apt.2015.12.004.

[27] Zheng, X.; Li, X.; Li, J.; Wang, L.; Jin, W.; Liu, J.; Pei, Y.; Tang, K. Efficient Removal of Anionic Dye (Congo Red) by Dialdehyde Microfibrillated Cellulose/Chitosan Composite Film with Significantly Improved Stability in Dye Solution. *Int. J. Biol. Macromol.* **2018**, *107*, 283–289. DOI: 10.1016/j.ijbiomac.2017.08.169.

[28] Mandal, S.; Calderon, J.; Marpu, S. B.; Omary, M. A.; Shi, S. Q. Mesoporous Activated Carbon as a Green Adsorbent for the Removal of Heavy Metals and Congo Red: Characterization, Adsorption Kinetics, and Isotherm Studies. *J. Contam. Hydrol.* **2021**, *243*, 103869. DOI: 10.1016/j.jconhyd.2021.103869.

[29] Han, D.; Zhao, H.; Gao, L.; Qin, Z.; Ma, J.; Han, Y.; Jiao, T. Preparation of Carboxymethyl Chitosan/Phytic Acid Composite Hydrogels for Rapid Dye Adsorption in Wastewater Treatment. *Colloid. Surf. A Physicochem. Eng. Asp.* **2021**, *628*, 127355. DOI: 10.1016/j.colsurfa.2021.127355.

[30] Ghorai, S.; Kumar, A.; Panda, A. B.; Pal, S. Effective Removal of Congo Red Dye from Aqueous Solution Using Modified Xanthan Gum/Silica Hybrid Nanocomposite as Adsorbent. *Bioresour. Technol.* **2013**, *144*, 485–491. DOI: 10.1016/j.biortech.2013.06.108.

[31] Saroj, S. K.; Pal, S.; Nagarajan, R. Polyol Intercalation in Copper Substituted Zinc Hydroxide Acetate and Evaluation of Its Adsorptive Role towards Congo Red Dye. *Appl. Clay Sci.* **2020**, *185*, 105411. DOI: 10.1016/j.clay.2019.105411.

[32] Miyata, S. The Synthesis of Hydrotalcite-Like Physico-Chemical Properties I : The Systems MgZ+-A13+-NO3, MgZ+-A13+-C1 -, Mg 2 +-A13 +-ClOg, Ni 2 +-A13 +-C1- and Zn 2 +-A13 +-C1. *Clays Clay Miner.* **1975**, *23*, 369–375. DOI: 10.1346/CCMN.1975.0230508

[33] Nagarajan, R.; Gupta, P.; Singh, P.; Chakraborty, P. An Ethylene Glycol Intercalated Monometallic Layered Double Hydroxide Based on Iron as an Efficient Bifunctional Catalyst.

Dalton Trans. **2016**, *45*, 17508–17520. DOI: 10.1039/c6dt03129c.

[34] Stanimirova, T.; Hibino, T. Ethylene Glycol Intercalation in MgAlCO 3 Hydrotalcite and Its Low-Temperature Intermediate Phases. *Appl. Clay Sci.* **2006**, *31*, 65–75. DOI: 10.1016/j.clay.2005.08.003.

[35] Tunney, J. J.; Detellier, C. Preparation and Characterization of Two Distinct Ethylene Glycol Derivatives of Kaolinite. *Clays Clay Miner.* **1994**, *42*, 552–560. DOI: 10.1346/CCMN.1994.0420506.

[36] Kasai, A.; Fujihara, S. Layered Single-Metal Hydroxide/Ethylene Glycol as a New Class of Hybrid Material. *Inorg. Chem.* **2006**, *45*, 415–418. DOI: 10.1021/ic051528d.

[37] Rasines, I.; García Casado, P. Thermal Decomposition of Cobalt and Nickel Double Hydroxy-Chlorides. *Thermochim. Acta* **1980**, *36*, 351–357. DOI: 10.1016/0040-6031(80)87030-4.

[38] Xi, Y.; Davis, R. J. Intercalation of Ethylene Glycol into Yttrium Hydroxide Layered Materials. *Inorg. Chem.* **2010**, *49*, 3888–3895. DOI: 10.1021/ic1000478.

[39] Wang, X.; Zhou, S.; Xing, W.; Yu, B.; Feng, X.; Song, L.; Hu, Y. Self-Assembly of Ni – Fe Layered Double Hydroxide/Graphene Hybrids for Reducing fi Re Hazard in Epoxy. *J. Mater. Chem. A* **2013**, *1*, 4383–4390. DOI: 10.1039/c3ta00035d.

[40] Yan, H.; Wang, J.; Zhang, Y.; Hu, W. Preparation and Inhibition Properties of Molybdate Intercalated ZnAlCe Layered Double Hydroxide. *J. Alloys Compd.* **2016**, *678*, 171–178. DOI: 10.1016/j.jallcom.2016.03.281.

[41] Mallakpour, S.; Azimi, F. *Spectroscopic Characterization Techniques for Layered Double Hydroxide Polymer Nanocomposites;* Elsevier Ltd.: Amsterdam, Netherlands, **2020**; pp. 231–280.

[42] Li, W.; Fan, G.; Yang, L.; Li, F. Surface Lewis Acid-Promoted Copper-Based Nanocatalysts for Highly Efficient and Chemoselective Hydrogenation of Citral to Unsaturated Allylic Alcohols. *Catal. Sci. Technol.* **2016**, *6*, 2337–2348. DOI: 10.1039/C5CY01084E.

[43] Stafford, K.; Sing, W. Reporting Physisorption Data for Gas/Solid Systems with Special Reference to the Determination of Surface Area and Porosity. *Pure Appl. Chem.* **1982**, *54*, 2201–2218. DOI: 10.1351/pac198254112201.

[44] Alothman, Z. A. A Review: Fundamental Aspects of Silicate Mesoporous Materials. *Materials (Basel)* **2012**, *5*, 2874–2902. DOI: 10.3390/ma5122874.

[45] Aouni, A.; Fersi, C.; Cuartas-Uribe, B.; Bes-Pía, A.; Alcaina-Miranda, M. I.; Dhahbi, M. Reactive Dyes Rejection and Textile Effluent Treatment Study Using Ultrafiltration and Nanofiltration Processes. *Desalination* **2012**, *297*, 87–96. DOI: 10.1016/j.desal.2012.04.022.

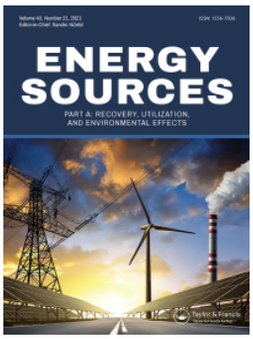
[46] Miyata, S. Anion-Exchange Properties of Hydrot Alcite-like Compounds. *Clays Clay Miner.* **1983**, *31*, 305–311. DOI: 10.1346/CCMN.1983.0310409.

[47] Zubair, M.; Daud, M.; McKay, G.; Shehzad, F.; Al-Harthi, M. A. Recent Progress in Layered Double Hydroxides (LDH)-Containing Hybrids as Adsorbents for Water Remediation. *Appl. Clay Sci.* **2017**, *143*, 279–292. DOI: 10.1016/j.clay.2017.04.002.

[48] Olalekan, A. P.; Olatunya, A.; Ekiti, A.; A.o, D. Langmuir, Freundlich, Temkin and Dubinin – Radushkevich Isotherms Studies of Equilibrium Sorption of Zn 2 + unto Phosphoric Acid Modified Rice Husk. *IOSR J. Appl. Chem.* **2012**, *3*, 38–45. DOI: 10.9790/5736-0313845.

[49] Ho, Y. S.; McKay, G, G.M. Pseudo-Second Order Model for Sorption Processes Y.S. *Process Biochem.* **1999**, *34*, 451–465. DOI: 10.1016/S0032-9592(98)00112-5.

[50] Wang, J.; Chen, C. Biosorbents for Heavy Metals Removal and Their Future. *Biotechnol. Adv.* **2009**, *27*, 195–226. DOI: 10.1016/j.biotechadv.2008.11.002.



Energy Sources, Part A: Recovery, Utilization, and Environmental Effects

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ueso20

Advanced Intelligent approach for state of charge estimation of lithium-ion battery

Deepak Kumar, M. Rizwan & Amrish K. Panwar

To cite this article: Deepak Kumar, M. Rizwan & Amrish K. Panwar (2023) Advanced Intelligent approach for state of charge estimation of lithium-ion battery, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 45:4, 10661-10681, DOI: 10.1080/15567036.2023.2249427

To link to this article: https://doi.org/10.1080/15567036.2023.2249427

Published online: 23 Aug 2023.

Submit your article to this journal

Article views: 34

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ueso20

ENERGY SOURCES, PART A: RECOVERY, UTILIZATION, AND ENVIRONMENTAL EFFECTS
2023, VOL. 45, NO. 4, 10661–10681
https://doi.org/10.1080/15567036.2023.2249427

Taylor & Francis
Taylor & Francis Group

Check for updates

Advanced Intelligent approach for state of charge estimation of lithium-ion battery

Deepak Kumar[a,b], M. Rizwan[a], and Amrish K. Panwar[b]

[a]Centre of Excellence for Electric Vehicle and Related Technologies, Department of Electrical Engineering, Delhi Technological University, Delhi, India; [b]Lithium-Ion Battery Technology Laboratory, Department of Applied Physics, Delhi Technological University, Delhi, India

ABSTRACT

The commercialization of lithium-ion batteries (LIBs) is rapidly increasing due to a variety of inherent and extrinsic parameters. The State of Charge (SOC), which denotes the amount of remaining capacity, is one of the most important performance metrics for these batteries. As a result, achieving a reliable and precise SOC estimation is essential for the greatest durability and security of LIBs. Estimating the SOC is important to improve the performance and robust utilization of LIBs. Here, this paper uses artificial neural network-based machine learning and deep learning approaches to estimate the battery state of charge. The battery voltage, current, and temperatures have been precisely integrated as input for the models. The proposed model's accuracy, reliability, and robustness are evaluated using available datasets. The mean absolute error was found to be in the range of 0.0030 to 0.0035, and root mean square errors 0.0043 to 0.0047 were obtained at 0 and 10°C operating temperatures. The outcomes demonstrate that the models can successfully estimate the SOC under different temperature conditions.

ARTICLE HISTORY
Received 1 June 2023
Revised 10 August 2023
Accepted 14 August 2023

KEYWORDS
Electric Vehicle; lithium-ion battery; state of charge; convolutional neural network; artificial intelligence

1. Introduction

The development of modern society is expected to perceive the evolution from nonrenewable fuel sources to renewable energy. Fossil fuel burning in vehicles releases pollutants including carbon monoxide, nitrogen oxides, and hydrocarbons into the atmosphere, which have been a major source of concern across the world (Kumar and Mehata 2022; Kumar, Singh, and Mehata 2022). Electric vehicle (EV) advancements are proliferating to reduce carbon dioxide (CO_2) emissions from traditional transportation (He et al. 2022). The worldwide EVs sales are over 26 million and stocks are rapidly increasing nearly five times more than in 2018 (Global EV Outlook 2023). The propulsion system of EVs is entirely electric, deriving power primarily from its onboard pack of batteries. Therefore, battery storage technology will become one of the most critical essential supporters of the green energy revolution (He et al. 2022). EVs commonly incorporate lithium-ion batteries (LIBs). These are most widely used as a power source for EVs because they deliver high energy density, long lifespan, lightweight, and low self-discharge rate. The sudden change in battery temperatures during over-charging/discharging, and other variables affect the safety and performance of LIBs (Li et al. 2022). The life span of the battery depends upon how it is utilized. The improper battery operations reduce battery life, create short circuits, and result in EVs catching fire. Therefore, battery monitoring is necessary since the battery system determines the vehicle's life, performance, and safety (Xiong et al. 2020). Increasing the driving range and maximizing battery runtime are critical factors in the modeling of EVs. The capability to monitor, regulate, and optimize the performance of single or numerous battery modules in an EV, and the ability to isolate modules from the system in the case of a failure and unexpected circumstances are

CONTACT M. Rizwan rizwan@dce.ac.in Centre of Excellence for Electric Vehicle and Related Technologies, Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India

© 2023 Taylor & Francis Group, LLC

crucial components of any EVs (Hossain Lipu et al. 2021). Therefore, the well-known battery management system (BMS) is a critical component for EVs. It can respond to both internal and external events of the battery and ensure a reliable and safe battery operation (Degla et al. 2023). In addition, it enhances battery performance while keeping necessary safety requirements ranges. One of the most important roles of BMS is to estimate the SOC (Tian et al., 2021).

The major challenge in battery state estimation research is SOC estimation, and it is also a most significant battery state. An accurate SOC estimation makes it possible to use LIBs packs to their highest capacity before it is removed, which significantly reduces the overall expenditure of production along with consumption. Furthermore, it is critical for ensuring the appropriate operation of the electric drive, although it effectively assesses a vehicle's outstanding driving range and is required for battery management and maintenance (McCurlie, Preindl, and Emadi 2017). It also prevents over-charging/discharging of the batteries, hence regarded as the crucial parameter of the BMS (Musiak and Li 2021). The SOC is defined as the remaining charge within the battery pack and is given by the remaining capacity of the battery divided by its available maximum capacity as represented in Equation (Eq.) 1 (Chemali et al. 2016).

$$\mathrm{SOC} = \frac{\mathrm{Q_r}}{\mathrm{Q_{max}}} \times 100\% \tag{1}$$

The remaining battery capacity represents by Q_r, while the maximum available capacity is Q_{max}. SOC is the % statistic representing the amount of available charge in the battery. These numbers stay between 0% and 100%, while they may exceed these boundaries in the case of an over-charge and discharge. Measuring SOC involves nonlinear function, fluctuations observed during rising temperature in charge and discharge of the battery (Cui et al. 2020). Regrettably, SOC cannot be measured directly because no method exists (Zou et al. 2015). Therefore, it is important to estimate SOC accurately. The model-based and data-driven are the two most common methods are used in SOC estimation techniques (Song et al. 2020). The few well-known model-based methods are, Electrochemical Impedance Model (Chang et al. 2022), Equivalent Circuit Model (Pai, Liu, and Ye 2023), Luenberger Observer (Ceraolo et al. 2020), Electrochemical Model (Y. F. Li et al. 2021), and Kalman filters (Dawei et al. 2022). These methods require deep knowledge of the electrical and electrochemical characteristics, and complex mathematical models of the LIBs used to model them (Liu et al. 2022). The family of Kalman filters attracted attention toward estimating SOC. This process delivers dependable and accurate models, but it requires in-depth domain expertise and careful engineering of features and take mostly prolonged development times (Xiong et al. 2023). These methods have inherent limitations, such as do not scale well across various battery cell chemistries. For example, the model construction must be updated due to changes in battery chemistry and required reparameterization of the models (Ma, Hu, and Cheng 2021). Furthermore, it also does not take into consideration cell irregularities such as unreliable manufacturing, irregular functioning issues, cell deterioration, and other factors (Tian et al. 2022). These issues were faced by the researchers. Therefore, the adoption of a data-driven method to estimate SOC is attractive and more popular among researchers.

In recent years, developments in artificial intelligence (AI) have accelerated the advance of data-driven SOC estimate methods (Tian et al. 2022). They do not use any physical modeling, instead, learn directly the mapping between SOC and measured data such as current, voltage, and temperatures. As a result, it eliminates the necessity of physical engineering of features, which can involve plenty of effort and time and require in-depth field knowledge (How et al., 2020). The literature has demonstrated excellent precision for such simple-to-implement methodologies, which avoid the difficulties imposed due to battery models (Wang et al. 2022). (Sheikh et al., 2023) Artificial neural networks (Hussein 2015), recurrent neural networks (Bian et al. 2022), support vector machine (Wu et al. 2022), extreme learning machines, Gaussian process regression (Fan, Wan, and Jiang 2022), optimized neural networks (Guo et al. 2021), and fuzzy logic (Singh, Fennie, and Reisner 2004) are frequently used in data-driven SOC estimation methods. Recently, recurrent neural networks (RNNs) were introduced to analyze sequence data, which are significant studies by researchers to address the dynamic profiles, to the directly measured battery SOC estimation (Tian, Xiong, and Shen 2019). In order to achieve a precise SOC estimate by applying voltage

and current sequences as inputs, two advanced RNN variants; long short-term memory (LSTM) and gated current unit (GRU) have been used. For instance, the authors applied a dataset with variable room temperature, which produced an average mean absolute error (MAE) of 2.17%, but it takes 40–50 hours to train the model having 85,000 epochs (Chemali et al. 2018). Yang et al. (Yang et al. 2019a) investigated SOC estimation directly by applying GRU using untrained temperatures and achieved root mean square (RMSE) of approximately 3.5% at the large number of epochs and several hours to train the models. Xiao et al. (Xiao, Liu, and Xiao 2019) also developed a deep convolutional model to reduce the training time of the model and divided the model training into two stages: pre-training and fine-training, while it took significant hours to get the desired results by keeping a few epochs and hidden neurons. Zhang et al. (Zhang et al. 2021) trained hybrid convolutional and RNNs models within a significant number of epochs, achieving an RMSE of under 2% at various temperature ranges, which consumed around 3.9 hours. Such studies required more time to achieve estimation accuracy. Furthermore, in circumstances when there is enough data available, it often takes several hours or days to finish training AI models. For example, we estimate SOC by implementing many hidden neurons and epochs, which consume more computational time, model training time, and high randomness of hyperparameters (Ren et al. 2021). Most AI models employ RNN architectures, which may be effective with sequence data like the SOC, but are challenging and time-consuming to train. Finally, even though RNN architecture effectively handles lengthy sequences, they are still vulnerable to vanishing gradients for longer sequences. The above existing AI-based approaches have research gaps within SOC estimation.

The main objective of this research work is to carry out, to minimize model training time, and to achieve higher accuracy in a short time. Furthermore, convolution neural networks (CNN) and RNNs such as LSTM and GRU are generally used in many fields, including computer vision and natural language processing, and battery state estimation. This article proposed LSTM, GRU, and CNN architecture for SOC estimation. Apart from that, it makes it possible to take minimum time and achieve higher estimation accuracy. Furthermore, we consistently achieved more accurate estimates of the SOC to contribute to effective BMS for EVs. The experimental platform uses an open-source library, TensorFlow, based on Python 6.4.12 and extensively employed in an artificial neural network. The contribution of this paper is as follows:

- Three AI-based models are proposed to capture the nonlinear dynamics of the battery under different temperature conditions. The model's best hyperparameters were selected using a step-by-step screening methodology.
- The battery statistics inside the selected window are connected to the "multi-time input" architecture that was developed to improve model input and SOC estimation performance. The impact of the battery statistics on the results may be effectively controlled to strengthen the estimated stability and accuracy.
- These approaches are model-free, no need to build OCV-SOC lookup tables for different temperatures and corresponding circuit types. The networks exhibit strong adaptability against unknown initial states, efficiently learn battery dynamics, and produce accurate SOC estimates at various temperatures.
- These models significantly reduce errors and maximize performance. The suggested methods contributed toward short computational times to obtain higher estimation accuracy. The presented approach should be adaptable to the SOC estimation on various batteries. These models may be applied in battery-related future work.

The rest of this paper is organized as follows: Section 2 introduces the proposed works and models mechanism architecture. Section 3 describes the experimental platform and data collection, processing model training, and the effect of temperature on LIBs. Section 4 verifies the proposed approach on variable temperatures with different initial and final SOC levels. Results and recommendations for further study are given in Section 5.

2. Proposed work

The entire process of the proposed work can be described as follows illustrated in Figure 1:
Step-1: Collect battery-tested data including current, voltage, capacity, and temperature measurements.
Step-2: Remove any outliers, smooth the data, and align all measurement data. Normalize the data by using a min-max scalar.
Step-3: Examine the data that have been processed and extract appropriate features such as the maximum voltage, current temperature, minimum voltage, current, and temperature. Split into training data.
Step-4: Train the deep learning model, i.e.,, CNN, GRU, LSTM. Hyperparameters tuning is done to provide the best performance of the model.
Step-5: Estimation of SOC output. The statistical performance of the models is calculated by calculating MAE, MSE, and RMSE.

2.1. Short-term long memory

The LSTM is a kind of RNNs commonly incorporated to learn long-term dependencies except when dealing with lost gradient difficulties (Srivastava and Lessmann 2018). It has the tendency to recall statistics for prolonged periods (Ko et al. 2021). The LSTM model's general architecture primarily contains input, output, and forget gates, as illustrated in Figure 2. The following activities are used to characterize an LSTM cell forward pass. In which input gate (i_k), forget gate (f_k), output-gate (o_k), hidden state (h_k), and cell memory (C_k) at time step k are each denoted, respectively.

The LSTM cell evaluates what statistics from cell state C_{k-1} determination is forgotten. The sigmoid function is executed by the forget-gate, which outputs are always between 0 and 1. The first step is determining which data will be lost from cell state C_{k-1}. The outputs of the sigmoid function implemented by the forget-gate are always between 0 and 1. The outcome denotes consequently how much should be forgotten. Although discarding (0) and keeping (1) everything representing, respectively, from the preceding cells state (Zhou, Feng, and Li 2021). A forget gates, described in Eq. (2) determines whether statistics must be removed from its cell memory (Bockrath et al. 2019).

$$f_k = \sigma(W_f * [h_{k-1}, X_k] + b_f) \tag{2}$$

Whereas sigmoid function (σ), defined as: $\sigma = \frac{1}{1+e^{-x}}$. The weight matrix ($W_f$, W_i, W_o), while the input vector ($X_k = V_k$, I_k, T_k) at time step k, and bias b_f, b_c, b_i, and b_o (Chemali et al. 2018a). The input gate recollects the new statistics into cell memory C_k. In Eq. (2, 3), whereas input gate is expressed Eq. (3).

$$i_k = \sigma(W_i * [h_{k-1}, X_k] + b_i) \tag{3}$$

$$_k = tanh\left(W_c * \left[h_(k-1), X_k\right] + b_c\right) \tag{4}$$

Whereas tan hyperbolic function, defined as: $tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.While "input gate i_k," formed by the sigmoid layer, decides which first value must be updated. Subsequently, the new vector value O_k, produced by a tanh layer that may be added to the states. The Eq. (2–4) outputs are now merged to update the unit memory C_k, as described by Eq. (5).

$$C_k = f_k * C_{k-1} + i_k * \hat{C}_k \tag{5}$$

The final output gate concept will also be discussed below. The output depends on the C_k. In Eq. (6), a sigmoid function was employed to finishing, which portion of C_k will be used,

$$O_k = \sigma(W_o * [h_{k-1}, X_k] + b_o) \tag{6}$$

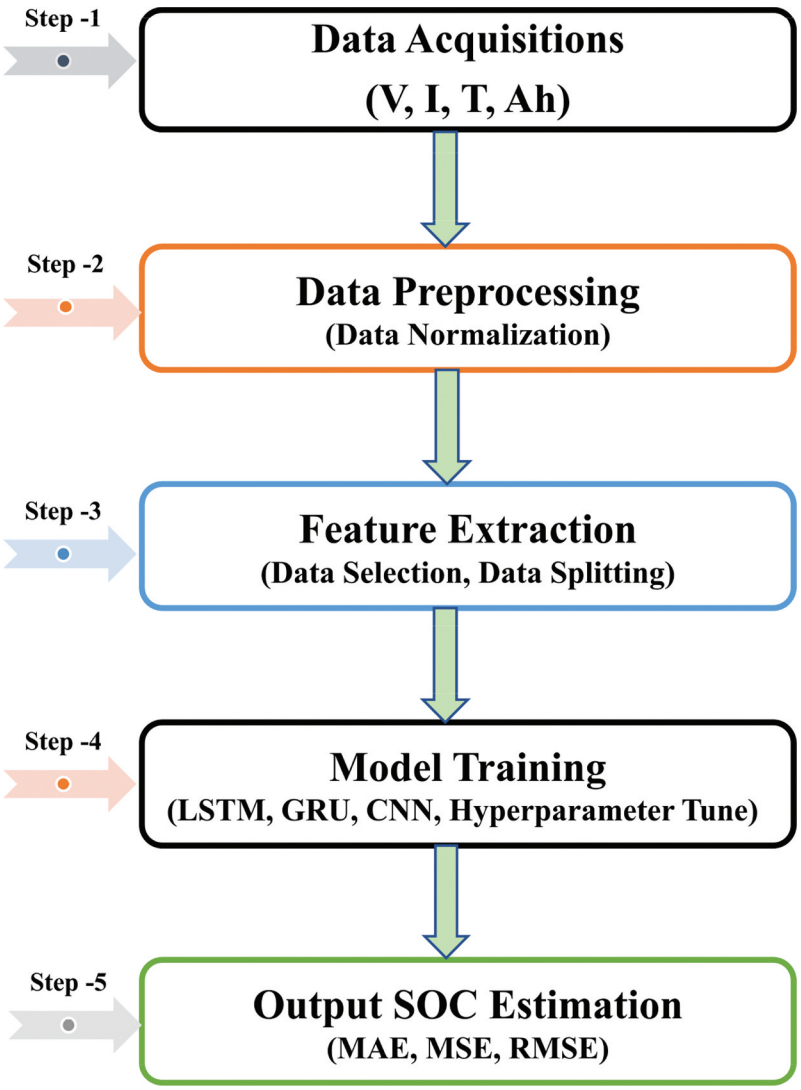


Figure 1. The flow chart of the proposed work.

Where output bias (b_0) is updated unit of states (Wang et al. 2019), the C_k is sent through a tanh function to generate the values from −1 to 1. After that, it is multiplied with the output $O_k$presented in Eq. (5) to produce the output h_k Eq. (7).

$$h_k = O_k * tanh(c_k) \tag{7}$$

While the output h_k. The fundamental drawback of the previous RNN was that earlier inputs vanished over time. Nevertheless, in the LSTM, input and forget gates regulate the cell memory. Thus, a long-term dependence issue is addressed (Naguib et al. 2021a). It is employed to comprehend the subtle complexity of intelligent behavior since it is the finest approach for modeling sequential data. The concept of "cell state" describes long-term memory. Due to their recursive structure, the previous information is stored in the cells.

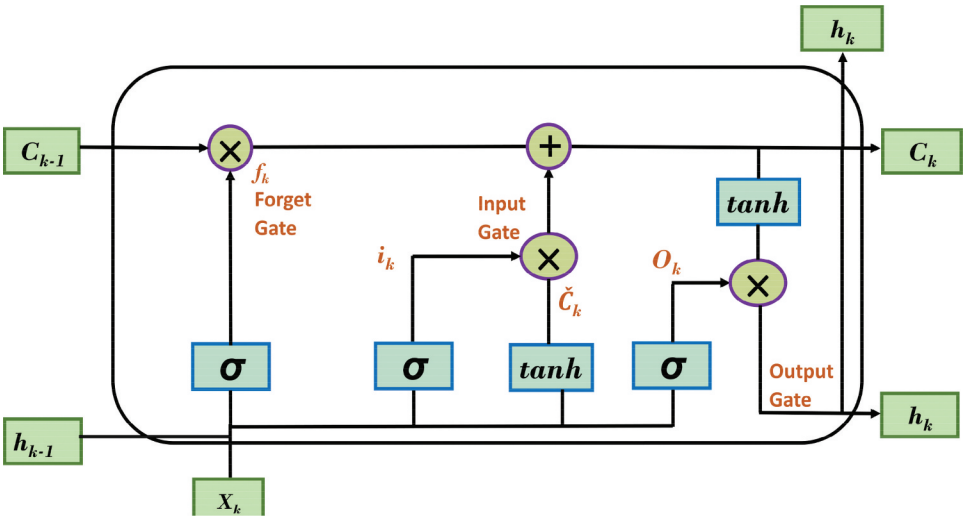


Figure 2. The internal architecture of a LSTM model.

2.2. Gated recurrent unit

An updated version of the LSTM termed GRU was introduced to leverage the connectivity established by a sequence of nodes to perform machine learning operations (Yang et al. 2019b). The fundamental difference between the LSTM and GRU; three gates in LSTM: input, forget, and output, whereas there are two gates in GRU: update and reset (Guo and Ma 2023). The GRU is less complex as compared to LSTM since it has fewer gates. In terms of accuracy, LSTM with three gates is meritorious, while it is a sophisticated model. While in GRU, the update gate governs how significantly preceding memory must be decided to save, and the reset gate regulates how to integrate a new input with previous inputs in its memory. The network acquires how its memory should operate by gaining knowledge of the configurations for its gates. The entire procedure of the GRU is illustrated in Figure 3.

In Equation (8–11), the reset gate "r_k" and the update gate "Z_k" are denoted (Zhao et al. 2017). The input at time k is X_k; the outputs (h_{k-1} and h_k) at times k-1 and k are, respectively. When reset gate weight is larger, then more data from the former neuron are vanished. To recollect important information in the neuron, increase the weight of the update gate. The update gate is represented by the Eq. (8) below, which defines how much of the GRU units is updated.

$$Z_k = \sigma(W_z * [h_{k-1}, X_k]) \tag{8}$$

Equation (8), generates the reset gate:

$$r_k = \sigma(W_r * [h_{k-1}, X_k]) \tag{9}$$

The reset gate hyperbolic tangent function is known as a new remember gate, and it is described by the following function.

$$\hat{h}_k = tanh(W_h[r_k * h_{k-1}, X_k]) \tag{10}$$

$$\hat{h}_k = (1 - z_k) * h_{k-1} + z_k * h_k \tag{11}$$

The weight matrices are W_z, W_r, W_h, and "×" denotes a multiplication of the corresponding matrix elements, requiring that the multiplication matrices be of the same type. The matrix adding operation is represented by the symbol "+." h_{k-1} typically relates to the retention of X_k and h_{k-1}.

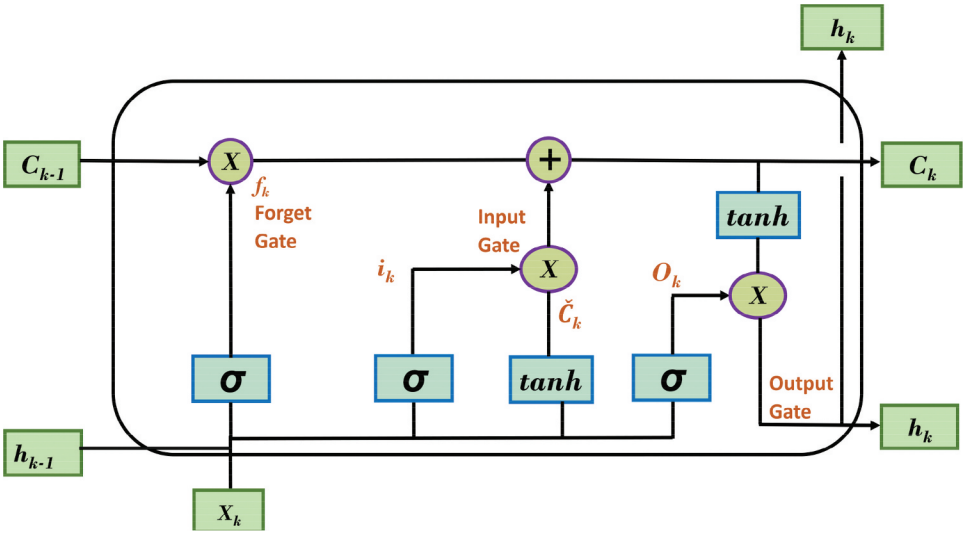


Figure 3. The internal architecture of a GRU model.

2.3. Convolutional neural network

In order to identify patterns and extract features, LeCun et al. introduced a convolutional neural network (CNN/ConvNet), a network of feedforward neurons (Lecun, Bengio, and Hinton 2015). One-dimensional (1D) CNN may identify the correlations among data input using convolutional layer filters and integrate them into more sophisticated feature parameters (Kiranyaz et al. 2021).

The regression classification is achieved using fully connected layers as well as additional neural networks. Furthermore, CNN is employed to analyze the battery datasets presented in this work, and the network topology is represented in Figure 4. Whereas CNN is acknowledged for its effectiveness with two-dimensional visuals, the same approach may be implemented for 1D data to extract superior characteristics, and numerous filters of the same size are implemented. The input parameters are convolved with every filter, and the results are merged as an output (Aman, Rizwan, and Kumar 2023). The following is the feature extraction equation for a time series using a one-dimensional convolution (Kuang and Xu 2018).

$$h_k = \sigma_{cnn}\left(W^*_{cnn}X_k + b_{cnn}\right) \tag{12}$$

Where the input signal "X_k," discrete convolution "*," filter weight "W_{cnn}," bias parameter "b_{cnn}" are learned throughout model training, while the underlying activation function "σcnn." The CNN maybe collect the spatial properties of the input using a few parameters and associating them to produce high-level topographies. After that, these characteristics are sent to the fully connected layer used for additional regression or classification. To capture the spatial properties of battery variables, 1D convolution is used in this research (Ren and Du 2023a). Furthermore, a convolutional layer was added, producing the dataset input to pass via activation function and convolutional operation before proceeding to the subsequent layer (Yao, Xu, and Ramezani 2021).

3. Experimental platform and data

3.1. Simulation platform

The experimental simulation was performed by the CPU (12th Gen Intel (R) Core (TM) i7–12700, 2.10 GHz, 12 core), Intel(R) UHD Graphics 770, and Windows 11 Pro operating system. The models were built by implementing an open-source library, TensorFlow, based on the

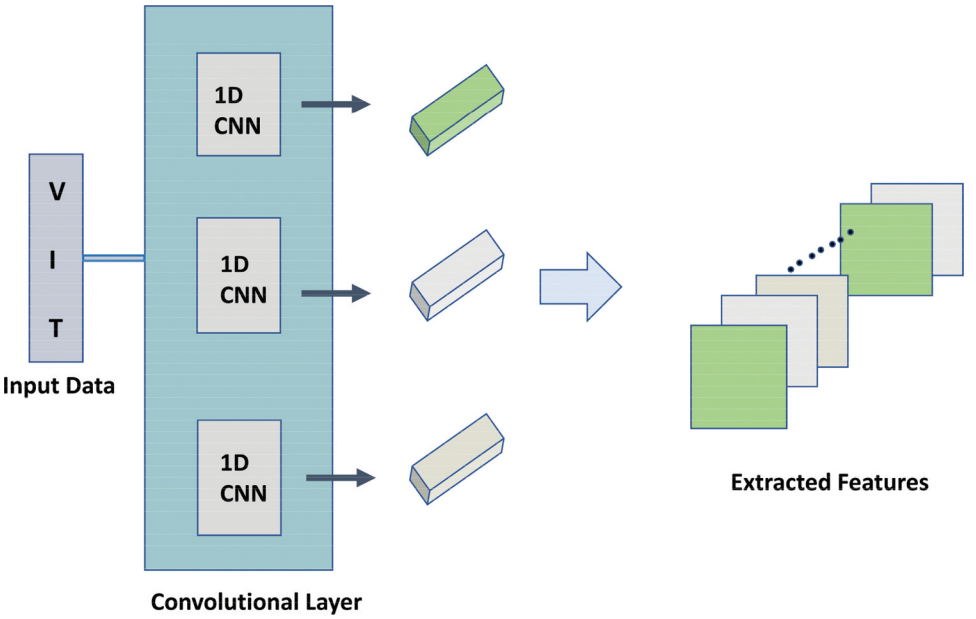


Figure 4. The structure of a 1D convolutional neural network.

Python 6.4.12 platform and extensively employed in an artificial neural network. The TensorFlow backend deep learning framework Keras and Python are the foundations for the neural network model. These models were introduced with dense layers, and window size with input data is 5 and 1, respectively. ReLu activation function, Adam optimization, and TensorFlow are used to output the layers' results (Peyal et al. 2021). Adam optimizer is used to stabilize the model's inherent learning rate. A linear activation function was incorporated with the output layer to get SOC as output. Subsequently, we map the learned states by dense layers to desired SOC output.

3.2. Data processing

This research was performed using the power cell's lithium-ion battery dataset: "Turnigy Graphene 5000mAh, 3S, 65C, LiPo, and Pack dimensions 144 × 51× 33 mm." The authors collected battery datasets from open-source publicly (Kollmeyer and Skells 2020). The battery characteristics are illustrated in Table 1. In this dataset, the test was performed in six different temperature environments (−20, −10, 0, 10, 25, and 40°C) with different C-rates (0.05C, 0.5C, 1C, and 2C) and four driving cycles, namely (UDDS, HWFET, LA92, and US06) of electric vehicles. Every 0.1 seconds, the test drive cycles were collected. In these test-driving cycles, we only used LA92 to perform SOC estimation. The SOC ranges from 0 to 1, but the voltage, current, and temperature measurements are highly variable. To increase precision and reliability, the measurement range must be scaled to the interval [0, 1]. To normalize each input data, the min-max normalization method is used in Eq.. 13.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{13}$$

Where x_{max} and x_{min} are the maximum and minimum data values, x represents the original data, and x_{norm} represents the normalized data. The normalized data are in the range [0, 1]. This study implements the "multi-time input" structure to window the input data for SOC estimation shown in Figure 5. The relationship between output and input strengthens the relationship

Table 1. Li-ion batteries characteristics.

Battery	Type
Model	Graphene 65C
Format	Pouch
Chemistry	LiPO or NCM
Rated Capacity	5.0 Ah
Normalized Resistance	14 mΩ
Specific Power and Energy	7.3 & 134k W/kg
Energy Density	329 Wh/l

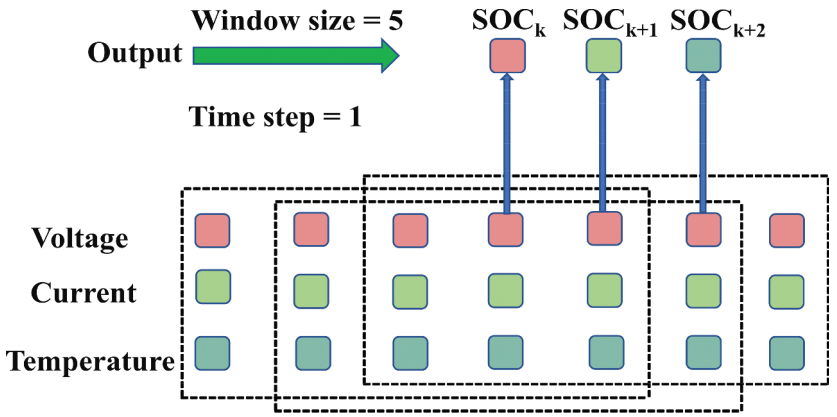


Figure 5. The "multi-time input" structure of the network.

between output and data in the window. All voltages, currents, and temperature measurements within that window are included in the input at time k. SOC_k applies to all data. This approach can increase the estimation accuracy and durability while maximizing the impact of all previous information on the outcomes (Naguib et al. 2021b).

3.3. Model training and hyperparameters tune

The Jupyter Notebook and Google Collab have used programming environments. The Keras library is used to implement the suggested models (Keras: The Python Deep Learning API). With the learning rate set to 0.01, the Adam algorithm is the selected optimizer to update the network weights and biases. All suggested models are trained for 50 epochs. The loss function is the root mean square error and mean squared error. The proposed network architectures, need the selection of different hyperparameter values to improve model accuracy and robustness. These values frequently have a significant impact on performances, if they are not set carefully. For example, an increase in computation time becomes large and costly, when the number of hidden is not selected carefully. Furthermore, the learning rate may be used to get accurate results and prevent overfitting and underfitting, while epochs are reducing the training load of networks (Ren and Du 2023b; Xie et al. 2023). In this article, all the hyperparameters are modified without traversing all possible standards, but rather by roughly setting different values of the parameters so that the training outcomes are within a satisfactory precision range.

3.4. Evaluation standard

The mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE), and maximum error (MaxError) are used to verify the estimation accuracy of the models calculated by Eq. (14–18).

$$MAE = \frac{1}{N}\sum_{k=1}^{N} \mid SOC_k - SOC_k^{'} \mid \tag{14}$$

$$MSE = \frac{1}{N}\sum_{k=1}^{N} \left(SOC_k - SOC_k^{'}\right)^2 \tag{15}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N} \left(SOC_k - SOC_k^{'}\right)^2} \tag{16}$$

$$R^2 = 1 - \frac{\Sigma_k \left(SOC_k^{'} - SOC_k\right)^2}{\Sigma_k \left(SOC_k - \overline{SOC}\right)^2} \tag{17}$$

$$MaxError = max \mid SOC_k - SOC_k^{'} \mid \tag{18}$$

Where "N" number of windows, "SOC_k" actual value, and $SOC_k^{'}$does the proposed network estimate the value. The arithmetic mean of the actual SOC value is $\overline{SOC}$. MAE evaluates how closely estimation is to actual values. The RMSE, on the other hand, indicates the variation of faults and is significantly more sensitive to considerable errors. The R^2 prediction coefficient is calculated as the assessment standard in order to highlight better the models' ability to fit data in this research. The network performs better the closer it is to 1.

3.5. Impact of temperatures on LIBs performance

The efficiency of a battery cell is highly reliant on its temperature; as a result, the operating temperature of the battery cell must constantly be maintained to stay within a specific range of temperature for battery safety, excellent performance, life cycle, and cycle stability (Jin et al. 2021). Therefore, the core problem is creating a battery energy storage technology that uses less energy and can more effectively avoid cell overheating. Operating circumstances, such as operation at high and low temperatures, have an impact on LIBs performance and shorten the battery's lifetime. When the temperature rises, it might result in thermal runaway, and the power and energy of the batteries are almost entirely lost (Divakaran et al. 2020). The variation in the working environment may have a significant impact on SOC estimation. The effect of dynamic load on SOC estimation, particularly under time-varying temperature conditions, has been addressed. The temperature change in the storage battery may inevitably occur during typical electric vehicle performance and will mainly affect SOC measurement (Li et al. 2021).

4. Analysis, simulation results, and performance comparison

In practice, battery loading varies significantly depending on how the battery is being used. For example, the battery-loading behaviors of hybrid vehicles while traveling on highways and in cities differs significantly with temperatures. Temperatures significantly influence the accuracy of SOC estimation. The suggested neural networks are trained based on AI approaches without needing a battery model. In this article, the authors trained the recommended neural network models using LA92 test drive cycle data at −10, 0, 10, 25, and 40°C to investigate the effect of temperature. To

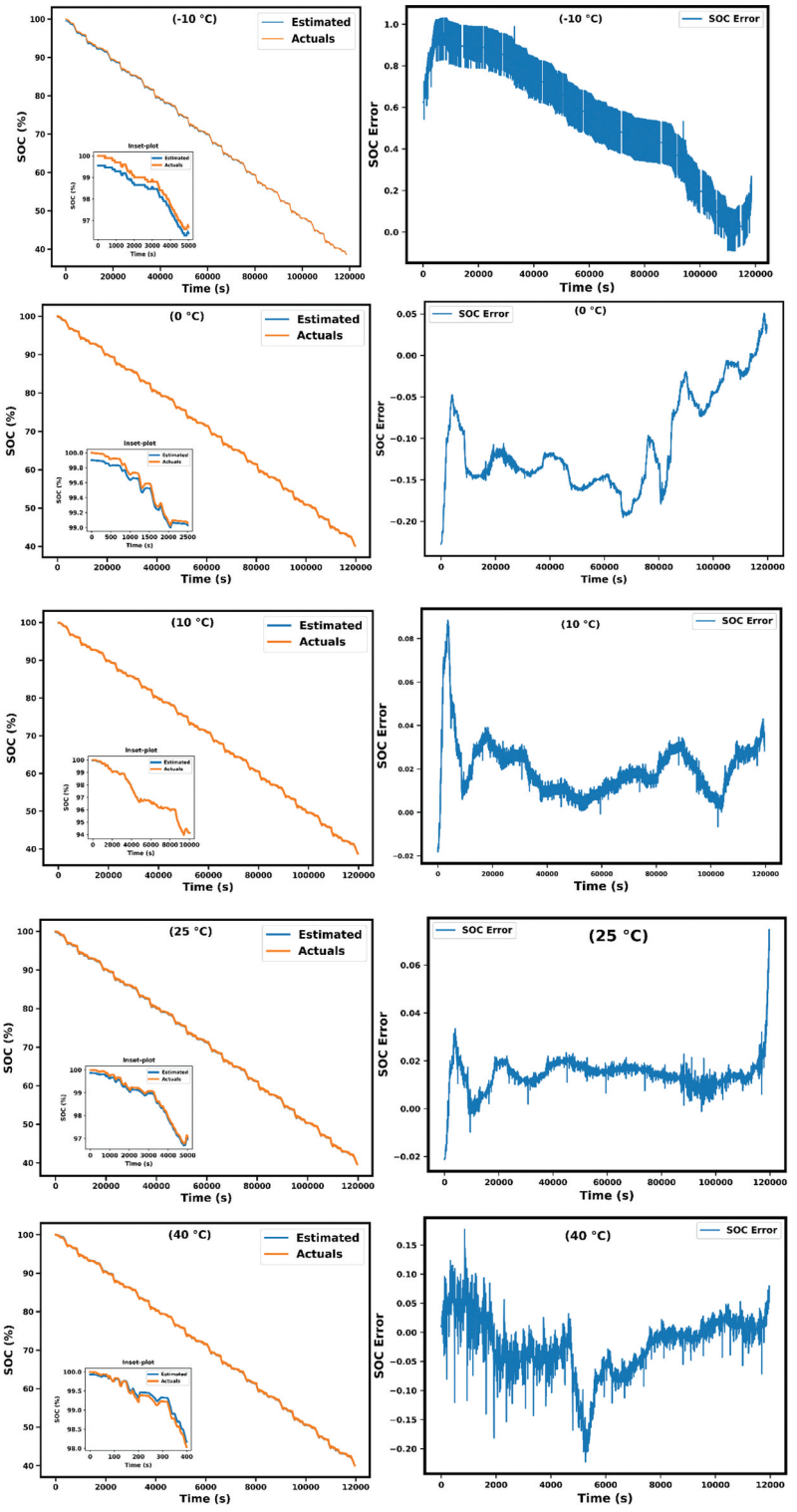


Figure 6. Actual and estimated performances and errors with LSTM under different temperatures.

Table 2. LSTM estimation accuracy results of LA92 test driving cycle at different temperatures.

S. No.	Temperature (°C)	MAE	MSE	RMSE	R^2	MaxError	Time (s)
1	−10	0.2202	0.0535	0.2314	0.9998	0.4710	425
2	0	0.1007	0.0169	0.1302	0.9999	0.3322	388
3	10	0.0279	0.0013	0.0368	0.9999	0.1063	393
4	25	0.0158	0.0003	0.0173	0.9999	0.0986	456
5	40	0.0174	0.0004	0.0220	0.9998	0.1037	60

Table 3. GRU estimation results of LA92 test driving cycle at different temperatures.

S. No.	Temperature (°C)	MAE	MSE	RMSE	R^2	MaxError	Time (s)
1	−10	0.2960	0.0934	0.3056	0.9997	0.5465	366
2	0	0.2328	0.0660	0.2569	0.9997	0.4982	392
3	10	1.0367	1.3256	1.1513	0.9956	2.0387	306
4	25	0.2615	0.0901	0.3003	0.9996	0.6111	351
5	40	0.2892	0.1165	0.3413	0.9995	0.7473	56

analyze the comparison between the applied algorithm and its performances on data inputs with temperatures. Additionally, we examine the comparability of the SOC estimation curves, and the errors of all datasets are uploaded into the standard LSTM, GRU, and CNN models. Models are run multiple times to decide the best parameters, such as optimizers, learning rates, epochs, etc.

Figure 6 shows estimated SOCs with errors using LSTM at different temperatures. It was found that at 25, 40, and 10°C, the LSTM model performed more efficiently, while very poor at −10 and 0°C with maximum error found to be 0.4710 and 0.3322, respectively. The LSTM accuracy outcomes are tabulated in Table 2. The high accuracy was obtained in ambient as well as high temperature. The fluctuations were observed at 0°C, as clearly seen in Figure 6. The zoom inset plot is also plotted in all figures. It might have occurred due to internal battery chemistry, which is very difficult to attribute at low temperatures. It was found that the model worked better with minimal errors, 0.0986 at ambient temperature and about 0.01037 at high temperature. Therefore, based on its maximum error, LSTM was found to be ineffective at low temperatures with this battery data because it is hard to know its internal chemistry at low temperatures.

The GRU model performs faster to train, taking small parameters, which is helpful in systems integration. To upload data into the GRU model to obtain estimates with actual results as tabulated in Table 3. At 10°C, performances of the GRU model are given notable fluctuations and significant errors between the estimated and actual SOC results. As seen in Figure 7, the fluctuations and errors become more effective at a temperature 10°C, which does not meet the accuracy and robustness requirements in the actual operation. It might be attributed that the performance of the GRU at 10°C does not give more accurate estimations. The main factor behind this is the more complicated internal temperature dynamics of lithium-ion batteries, which shows poor performance with variations in temperatures, which makes it considerably more difficult for the suggested network to get meaningful attributes. The GRU estimation performance at 0°C shows the best accuracy, RMSE, and Max error around 0.2569 and 0.4982, respectively. While at −10 and 40°C showing more errors. As a result, since temperatures significantly impact GRU networks, it is impossible to determine battery SOC at various temperatures. The estimated effectiveness of the model network is more precisely investigated using MAE, RMSE, and R^2, as shown in Table 3.

The CNN model of SOC estimation results with errors are shown in Figure 8. The estimated performance under various temperatures is generally satisfactory compared to LSTM and GRU. The SOC estimation performances, such as MAE, MSE, RMSE, $R^{2,,}$ and Max error tabulated in Table 4. The superior performance was gained by CNN at 0 and 10°C, while the performance was weakest at 25 and 40°C. The SOC estimate performance at −10°C is satisfactory compared to the ambient and high temperatures. It is significantly more difficult for the suggested network to gain relevant characteristics since lithium-ion batteries perform poorly and have more complicated internal dynamics at low temperatures. In brief,

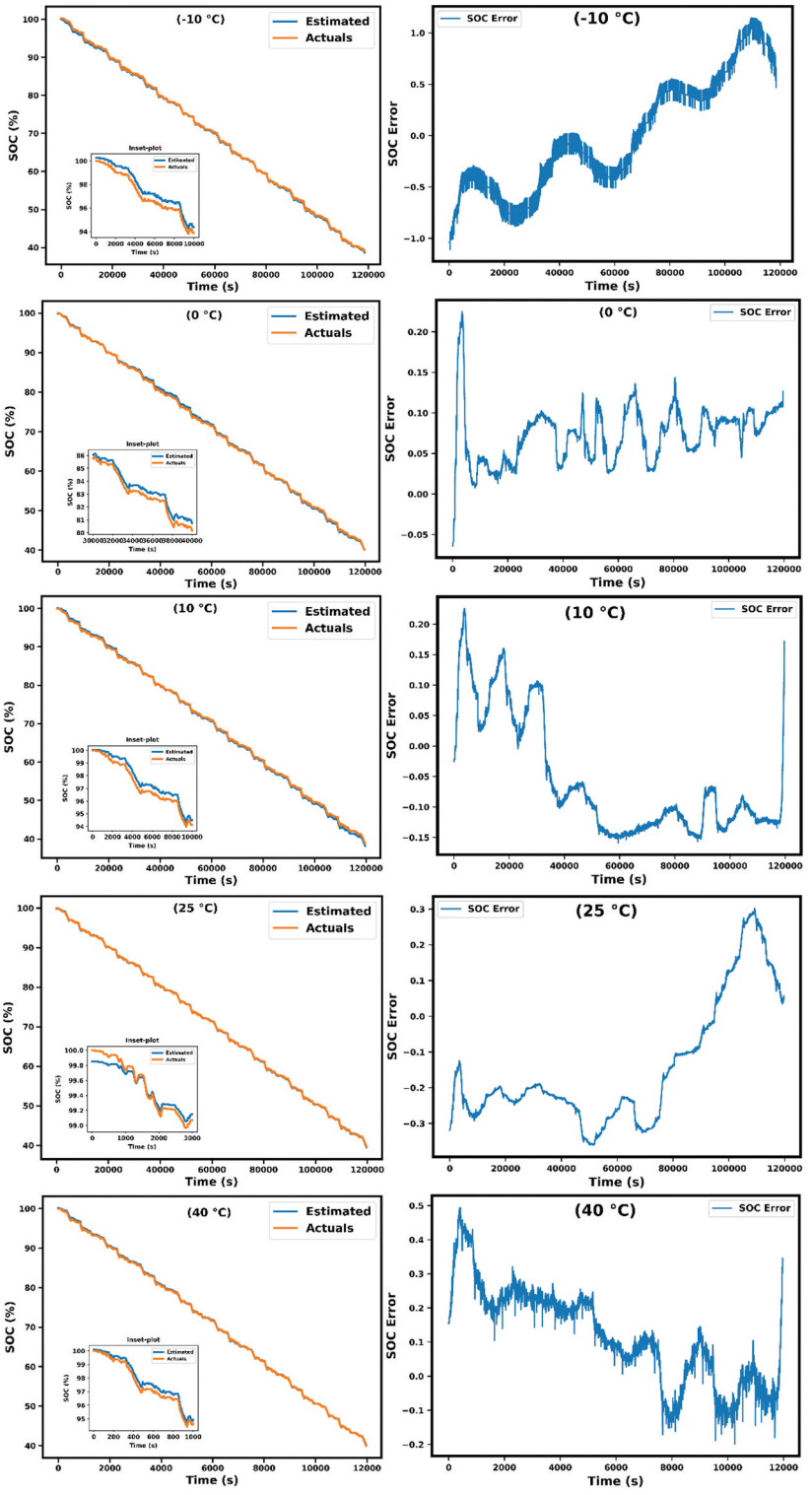


Figure 7. Actual and estimated performances and errors with GRU under different temperatures.

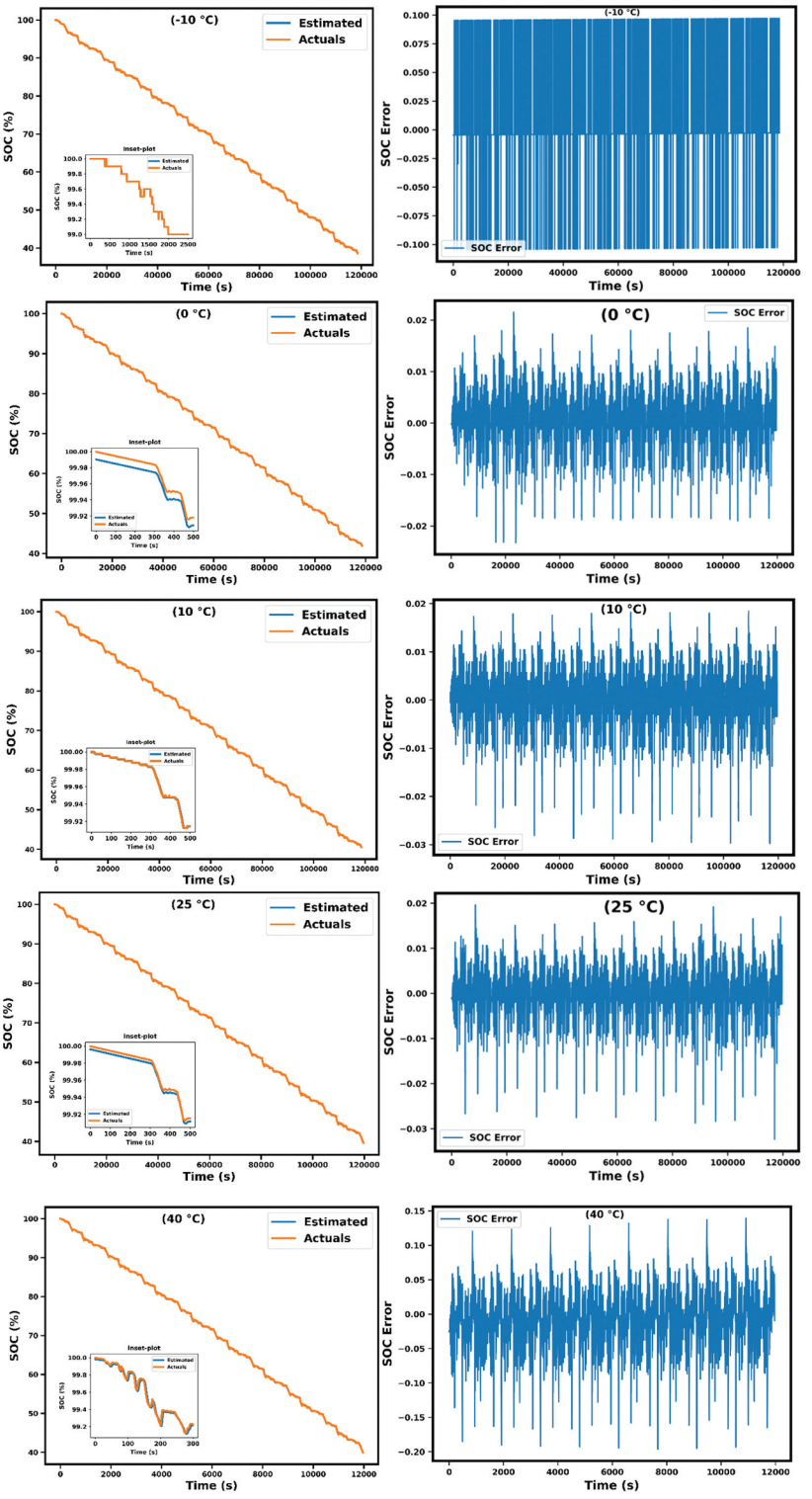


Figure 8. Actual and estimated performances and errors with CNN under different temperatures.

Table 4. CNN estimation results of LA92 test driving cycle at different temperatures.

S. No.	Temperature (°C)	MAE	MSE	RMSE	R^2	Max Error	Time (s)
1	−10	0.0033	0.0002	0.0148	0.9999	0.1007	141
2	0	0.0030	$1.85e^{-5}$	0.0043	0.9999	0.0215	243
3	10	0.0035	$2.29e^{-5}$	0.0047	0.9999	0.0286	218
4	25	0.0247	0.0006	0.0250	0.9999	0.0608	248
5	40	0.0310	0.0016	0.0403	0.9999	0.2069	33

Table 5. CNN estimation results of LA92 test driving cycle at different temperatures.

	Actual Computational Time (s)			Computational Efficiency (%)	
Temperature (°C)	LSTM	GRU	CNN	GRU	CNN
−10	425	366	141	13.88	66.82
0	392	388	243	1.03	37.36
10	393	306	218	22.14	44.53
25	456	351	248	23.03	45.61
40	60	56	33	6.67	45

the proposed network may easily be expanded to consider the impact of ambient temperature on battery SOC estimates by simply adding temperature as a network input. Overall, the suggested approach may provide a reasonable SOC estimate for various temperatures.

To validate the effectiveness of the proposed SOC estimation models LSTM, GRU, and CNN. CNN shows the best results and the worst performance achieved by GRU. Whereas GRU and LSTM take more training time due to their complex structure and many cell states present in each hidden neuron, CNN requires less time to train the model. This research demonstrates that the suggested network completely satisfies the requirements of practical operation and has excellent accuracy and resilience under various working circumstances. A comparison of different reports based on AI models for SOC estimation performance is tabulated in Table 6. In this comparison, the present work is more accurate in estimating SOC, demonstrating that the proposed models could be promising for more accurate SOC estimation. The computational efficiency (CE) of GRU and CNN method is calculated by using Eq. 19, by taking LSTM actual computational time as reference as shown in Table 5. The highest CE shown by CNN is 66.82% at −10°C, while lowest CE shown by GRU is 1.03% at 0°C.

$$\textit{Computational Efficiency}(\%) = \frac{\text{Reference computational time} - \text{Actual computational time}}{\text{Reference computational time}} \times 100 \qquad (19)$$

5. Conclusions

In EVs, the BMS monitors and regulates battery energy storage systems. Estimating SOC is the most essential function of a BMS. The accurate estimation of SOC ensures the safe and stable operation of LIBs in EVs, therefore ensuring the safety of drivers and vehicles. It is a compulsory task to redress the malfunctioning of batteries and improve battery lifespan. Moreover, it reduces the number of batteries needed in the battery pack, which helps lower costs by preventing LIBs from being overcharged or undercharged. The AI-based advancements have confirmed their acceptability in this domain. This paper proposed the LSTM, GRU, and CNN models for SOC estimation at different temperatures with the test drive cycle datasets. These models have the ability to learn to overcome the complexity and uncertainty during the estimation process. The results demonstrate that CNN models perform superior SOC estimation with RMSE about 0.0043 in approximately 243 seconds, while LSTM gives RMSE about 0.0173 in 456 seconds and GRU revealed 0.3003 in 351 seconds. The total time taken by LSTM, GRU, and CNN to perform all the experiments is 1722, 1471, and 883 in seconds, respectively. The CNN and GRU give results faster approximately 67% and 23% as compared to LSTM. Therefore, CNN is the best among these and saves time, which is very crucial for the BMS to take

Table 6. Comparison of accuracy performances of different reports for SOC estimation.

Algorithm	Input features & output	Configuration	Battery type/ Temperature (°C)	Performance Accuracy	Research gap	References
UKF	OCV–SOC–temperature table used	—-	LiFePO4/0–50	RMSE= 5	The complex calculation, time consuming	(Xing et al. 2014)
Adaptive neuro-fuzzy inference systems (ANFIS)	OCV and temperatures	ANFIS layers- 5 Epochs= 10	Lithium-ion cells/-30–55	MSE =5.6	Computational, interpret problems, time-consuming	(Awadallah and Venkatesh 2016)
PSO-LSTM	Input = [V, I, T]Output = [SOC] Output = [SOC]	hidden neurons – 150,400 Epochs – 200,700	LiFePO4/25	MAE= .3493	More hyperparameters randomness	(X. Ren et al. 2021)
LSTM	Input = [V, I, T]Output = [SOC] Output = [SOC]	hidden layers – 1 hidden neurons –500	LiNiCoAlO2/0, 10, 25	MAE = 1.606	More hyperparameters randomness	(Chemali et al. 2018b)
RNN-GRU	Input = [V, I, T]Output = [SOC] Output = [SOC]	hidden layers – 2 hidden neurons – 200 Epochs – 500	LiMn2O4/.5, 2, 5, 25, 45 /0.5, 2, 5, 25, 45	MSE = 2.99	Time consuming, network structure must be enhanced	(S. Li et al. 2021)
1D-CNN	Input = [V, I, T]Output = [SOC] Output = [SOC]	Convolution layer – 2 Epoch – 50	Turnigy Graphene–10, 0, 10, 25, 40 –10, 0, 10, 25, 40	MAE = 1.85e –5RMSE = .0043 RMSE = 0.0043		Present works

decisions earliest. In addition, it has the capability to estimate SOC much more accurately than other models. It can also learn the dynamic behavior of batteries more effectively. This work is expected to be a reference for researchers and industrialists to understand the SOC estimation process with the dynamic behavior of batteries. This proposed work may be extended to the future as:

- The battery pack will be used instead of the individual cells for state estimation on different dynamic loading profiles and temperatures.
- The different AI-based algorithms also will be joined together to contribute toward the advanced battery SOC estimate of the battery packs and verify the effectiveness of the proposed algorithm.
- The state estimate may include State of Health (SOH), Remaining Useful Life (RUL), and End of Life (EOL).

Abbreviations

Q_r	remaining battery capacity
Q_{max}	maximum available capacity
i_k	input gate
f_k	forget gate
o_k	output-gate
h_k	hidden state
C_k	cell memory
C_{k-1}	cell state
σ	sigmoid function
W_f	forget weight matrix
W_i	input weight matrix
W_o	output weight matrix
k	time step
X_k	input vector
$_{bf}$	forget gate bias
$_{bc}$	cell memory
$_{bi}$	input gate bias
b_o	output gate bias
tanh	tan hyperbolic function
r_k	reset gate
Z_k	update gate
W_z	update gate weight matrices
W_r	reset gate weight matrices
W_h	output weight matrices
"×"	multiplication of the corresponding matrix elements
"+"	matrix adding operation
X_k	input signal
*	discrete convolution
W_{cnn}	filter weight
b_{cnn}	bias parameter
σcnn	activation function
x_{max}	maximum data values
x_{min}	minimum data values
x	represents the original data
x_{norm}	represents the normalized data
LIBs	lithium-ion batteries
SOC	state of Charge

EV	Electric vehicle
LSTM	long short-term memory
GRU	gated current unit
CNN	convolution neural networks
BMS	battery management system
MAE	mean absolute error
MSE	mean squared error
RMSE	root mean square error
MaxError	maximum error
SOC_k	actual value
$SOC_k^{'}$	estimated value

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

No specific grant has been given to this research by funding agencies/organizations.

Notes on contributors

Deepak Kumar received his B.Tech. degree in Electrical and Electronics Engineering from Dr. A.P.J. Abdul Kalam Technical University Uttar Pradesh, Lucknow, India, in 2017, and M.Tech. degree in Nanoscience and Technology from the Department of Applied Physics at Delhi Technological University, Delhi, India, in 2021. He is currently pursuing PhD degree at Delhi Technological University, Delhi, India with the Department of Electrical Engineering. He has published two research articles in reputed international/national journals. His research interests include battery technology, energy storage in electric vehicles, and AI-based battery management ystems/battery state estimation.

Dr. M. Rizwan did his post-doctoral research at Virginia Polytechnic Institute and State University, USA. He is presently working as a Professor at the Department of Electrical Engineering, Delhi Technological University, Delhi, India. He has almost 20 years of teaching and research experience. Dr. Rizwan has successfully completed three research projects in the area of renewable energy systems and published and presented more than 175 research papers in reputed international/national journals including IEEE transactions and conference proceedings. Dr. Rizwan has published one book titled Grid Integration of Solar Photovoltaic Systems, CRC Press, (Taylor and Francis Group), 2017, and Edited one book for AIP Publishing, USA. He is the recipient of Raman Fellowships for Post-Doctoral Research for Indian Scholars in USA, DST Start up Grants (Young Scientists) and many more. His area of interest includes soft computing applications in power engineering, renewable energy systems, building energy management, smart grid etc. He is a Sr. Member of IEEE, Life Member of ISTE, Life Member of SSI, Member of International Association of Engineers (IAENG), and many other reputed societies.

Dr. Amrish K Panwar completed his MS degree from Chaudhary Charan Singh (CCS) University, India, and his PhD degree from Indian Institute of Technology (IIT) Roorkee, India in 2005. After completing PhD degree, he has worked as a post-doctoral research associate at the Department of Metallurgical and Materials Engineering, IIT Kharagpur, India. Before joining Delhi Technological University (DTU), he has also worked as senior lecturer in Jaypee University of Engineering and Technology (JUET), India. Presently, he is working as Assistant Professor in the Department of Applied Physics, Delhi Technological University, New Delhi, India. His area of research includes: energy storage and conversion devices (Li+ /Na+ batteries/supercapacitors/solar cells), surface modification, wetting adhesion, and nanocomposites.

ORCID

M. Rizwan http://orcid.org/0000-0003-2839-7241

Dataset availability statement

Dataset analysis and forecasting have been performed utilizing publicly available data from Mendeley Data.

Declaration

The authors declare that the present research has no financial or personal ties to any individual or group that could unreasonably affect this work. This manuscript is the author's work and does not contain material from other sources. All data measures are accurate and unaltered results, and this work has never been published or submitted to another journal for publication consideration. Also, none of the authors has any financial or scientific conflicts of interest concerning the research described in this manuscript.

References

Aman, R., M. Rizwan, and A. Kumar. 2023. Fault classification using deep learning based model and impact of dust accumulation on solar photovoltaic modules. *Energy Sources, Part A: Recovery, Utilization, & Environmental Effects* 45 (2):4633–51. doi:10.1080/15567036.2023.2205859.

Awadallah, M. A., and B. Venkatesh. 2016. Accuracy improvement of SOC estimation in lithium-ion batteries. *Journal of Energy Storage* 6:95–104. doi:10.1016/J.EST.2016.03.003.

Bian, C., S. Yang, J. Liu, and E. Zio. 2022. Robust state-of-charge estimation of Li-ion batteries based on multichannel convolutional and bidirectional recurrent neural networks. *Applied Soft Computing* 116:108401. doi:10.1016/J.ASOC.2021.108401.

Bockrath, S., A. Rosskopf, S. Koffel, S. Waldhor, K. Srivastava, and V. R. H. Lorentz (2019). State of charge estimation using recurrent neural networks with long short-term memory for lithium-ion batteries. *IECON Proceedings (Industrial Electronics Conference)*, 2019 October, 2507–11. 10.1109/IECON.2019.8926815

Ceraolo, M., G. Lutzemberger, D. Poli, and C. Scarpelli. 2020. Luenberger-based state-of-charge evaluation and experimental validation with lithium cells. *Journal of Energy Storage* 30:101534. doi:10.1016/J.EST.2020.101534.

Chang, C., S. Wang, C. Tao, J. Jiang, Y. Jiang, and L. Wang. 2022. An improvement of equivalent circuit model for state of health estimation of lithium-ion batteries based on mid-frequency and low-frequency electrochemical impedance spectroscopy. *Measurement* 202:111795. doi:10.1016/J.MEASUREMENT.2022.111795.

Chemali, E., P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi. 2018a. Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. *IEEE Transactions on Industrial Electronics* 65 (8):6730–39. doi:10.1109/TIE.2017.2787586.

Chemali, E., P. J. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi. 2018b. Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. *IEEE Transactions on Industrial Electronics* 65 (8):6730–39. doi:10.1109/TIE.2017.2787586.

Chemali, E., P. J. Kollmeyer, M. Preindl, and A. Emadi. 2018. State-of-charge estimation of Li-ion batteries using deep neural networks: a machine learning approach. *Journal of Power Sources* 400:242–55. doi:10.1016/J.JPOWSOUR.2018.06.104.

Chemali, E., M. Preindl, P. Malysz, and A. Emadi. 2016. Electrochemical and electrostatic energy storage and management systems for electric drive vehicles: state-of-the-art review and future trends. *IEEE Journal of Emerging and Selected Topics in Power Electronics* 4 (3):1117–34. doi:10.1109/JESTPE.2016.2566583.

Cui, S., X. Yong, S. Kim, S. Hong, and I. Joe. 2020. An LSTM-Based encoder-decoder model for state-of-charge estimation of lithium-ion batteries. *Advances in Intelligent Systems & Computing* 178–88. doi:10.1007/978-3-030-51965-0_15/COVER.

Dawei, Q., L. Zixuan, Y. Fan, F. Luyan, Z. Mingyue, and L. Haoxuan. 2022. State of charge estimation for the vanadium redox flow battery based on extended Kalman filter using modified parameter identification. *Energy Sources, Part A: Recovery, Utilization, & Environmental Effects* 44 (4):9747–63. doi:https://doi.org/10.1080/15567036.2022.2136800.

Degla, A., M. Chikh, M. Mzir, and Y. Belabed. 2023. State of charge estimation for Li-ion battery based intelligent algorithms. *Electrical Engineering* 105 (2):1179–97. doi:10.1007/S00202-022-01728-9/FIGURES/13.

Divakaran, A. M., D. Hamilton, K. N. Manjunatha, and M. Minakshi. 2020. Design, development and thermal analysis of reusable li-ion battery module for future mobile and stationary applications. *Energies* 13 (6):1477. doi:10.3390/EN13061477.

Executive summary – Global EV Outlook 2023 – Analysis. IEA. https://www.iea.org/reports/global-ev-outlook-2023/executive-summary.

Fan, K., Y. Wan, and B. Jiang. 2022. State-of-charge dependent equivalent circuit model identification for batteries using sparse Gaussian process regression. *Journal of Process Control* 112:1–11. doi:10.1016/J.JPROCONT.2021.12.012.

Guo, S., and L. Ma. 2023. A comparative study of different deep learning algorithms for lithium-ion batteries on state-of-charge estimation. *Energy* 263:125872. doi:10.1016/J.ENERGY.2022.125872.

Guo, Y., Z. Yang, K. Liu, Y. Zhang, and W. Feng. 2021. A compact and optimized neural network approach for battery state-of-charge estimation of energy storage system. *Energy* 219:119529. doi:10.1016/J.ENERGY.2020.119529.

He, H., F. Sun, Z. Wang, C. Lin, C. Zhang, R. Xiong, J. Deng, X. Zhu, P. Xie, S. Zhang, et al. 2022. China's battery electric vehicles lead the world: Achievements in technology system architecture and technological breakthroughs. *Green Energy and Intelligent Transportation* 1 (1):100020. doi:10.1016/J.GEITS.2022.100020.

Hossain Lipu, M. S., M. A. Hannan, T. F. Karim, A. Hussain, M. H. M. Saad, A. Ayob, M. S. Miah, and T. M. Indra Mahlia. 2021. Intelligent algorithms and control strategies for battery management system in electric vehicles: progress, challenges and future outlook. *Journal of Cleaner Production* 292:126044. doi:10.1016/J.JCLEPRO.2021.126044.

How, D. N. T., M. A. Hannan, M. S. H. Lipu, K. S. M. Sahari, P. J. Ker, and K. M. Muttaqi. 2020. State-of-Charge Estimation of Li-Ion Battery in Electric Vehicles: A Deep Neural Network Approach. *IEEE Transactions on Industry Applications* 56 (5):5565–74. doi:10.1109/TIA.2020.3004294.

Hussein, A. A. 2015. Capacity fade estimation in electric vehicle Li-ion batteries using artificial neural networks. *IEEE Transactions on Industry Applications* 51 (3):2321–30. doi:10.1109/TIA.2014.2365152.

Jin, L., J. Tian, S. Gao, P. Xie, M. Akbarzadeh, T. Kalogiannis, M. Berecibar, Y. Lan, D. Hu, Y. Ding, et al. 2021. A novel hybrid thermal management approach towards high-voltage battery pack for electric vehicles. *Energy Conversion and Management* 247:114676. doi:10.1016/J.ENCONMAN.2021.114676.

Kiranyaz, S., O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman. 2021. 1D convolutional neural networks and applications: asurvey. *Mechanical Systems and Signal Processing* 151:107398. doi:10.1016/J.YMSSP.2020.107398.

Ko, M. S., K. Lee, J. K. Kim, C. W. Hong, Z. Y. Dong, and K. Hur. 2021. Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind Power forecasting. *IEEE Transactions on Sustainable Energy* 12 (2):1321–35. doi:10.1109/TSTE.2020.3043884.

Kollmeyer, P., and M. Skells. 2020. Turnigy Graphene 5000mAh 65C Li-ion battery data. 1. doi:10.17632/4FX8CJPRXM.1.

Kuang, D., and B. Xu. 2018. Predicting kinetic triplets using a 1d convolutional neural network. *Thermochimica Acta* 669:8–15. doi:10.1016/J.TCA.2018.08.024.

Kumar, D., and M. S. Mehata. 2022. Synthesis of diamagnetic ZnO nano-crystallites via sol-gel method and their photocatalytic activity. *Indian Journal of Engineering & Materials Sciences* 29 (4):437–44. doi:10.56042/IJEMS.V29I4.48695.

Kumar, D., M. K. Singh, and M. S. Mehata. 2022. Exploration of grown cobalt-doped zinc oxide nanoparticles and photo-degradation of industrial dye. *Materials Research Bulletin* 150:150. doi:10.1016/J.MATERRESBULL.2022.111795.

Lecun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. Nature 521 (7553):436–44. doi:10.1038/nature14539.

Liu, S., X. Dong, X. Yu, X. Ren, J. Zhang, and R. Zhu. 2022. A method for state of charge and state of health estimation of lithium-ion battery based on adaptive unscented Kalman filter. *Energy Reports* 8:426–36. doi:10.1016/J.EGYR.2022.09.093.

Li, L., C. Wang, S. Yan, and W. Zhao. 2021. A combination state of charge estimation method for ternary polymer lithium battery considering temperature influence. *Journal of Power Sources* 484:229204. doi:10.1016/J.JPOWSOUR.2020.229204.

Li, S., C. Zhang, J. Du, X. Cong, L. Zhang, Y. Jiang, and L. Wang. 2022. Fault diagnosis for lithium-ion batteries in electric vehicles based on signal decomposition and two-dimensional feature clustering. *Green Energy and Intelligent Transportation* 1 (1):100009. doi:10.1016/J.GEITS.2022.100009.

Ma, L., C. Hu, and F. Cheng. 2021. State of charge and state of Energy estimation for lithium-ion batteries based on a long short-term memory neural network. *Journal of Energy Storage* 37:102440. doi:10.1016/J.EST.2021.102440.

McCurlie, L., M. Preindl, and A. Emadi. 2017. Fast model predictive control for redistributive lithium-ion battery balancing. *IEEE Transactions on Industrial Electronics* 64 (2):1350–57. doi:10.1109/TIE.2016.2611488.

Musiak, M., and Z. Li (2021). Real time ultrasonic monitoring of solid-state lithium-ion cells in the frequency domain. *2021 IEEE International Conference on Prognostics and Health Management, ICPHM 2021*. 10.1109/ICPHM51084.2021.9486632

Naguib, M., P. Kollmeyer, C. Vidal, and A. Emadi (2021a). Accurate surface temperature estimation of lithium-ion batteries using feedforward and recurrent artificial neural networks. *2021 IEEE Transportation Electrification Conference & Expo (ITEC)*, 52–57. 10.1109/ITEC51675.2021.9490043

Naguib, M., P. Kollmeyer, C. Vidal, and A. Emadi (2021b). Accurate surface temperature estimation of lithium-ion batteries using feedforward and recurrent artificial neural networks. *2021 IEEE Transportation Electrification Conference and Expo, ITEC 2021*, 52–57. 10.1109/ITEC51675.2021.9490043

Pai, H. Y., Y. H. Liu, and S. P. Ye. 2023. Online estimation of lithium-ion battery equivalent circuit model parameters and state of charge using time-domain assisted decoupled recursive least squares technique. *Journal of Energy Storage* 62:106901. doi:10.1016/J.EST.2023.106901.

Peyal, H. I., S. M. Shahriar, A. Sultana, I. Jahan, and M. H. Mondol (2021). Detection of tomato leaf diseases using transfer learning architectures: a comparative analysis. *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0, ACMI 2021*. 10.1109/ACMI53878.2021.9528199

Ren, L., J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen. 2021. A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. *IEEE Transactions on Industrial Informatics* 17 (5):3478–87. doi:10.1109/TII.2020.3008223.

Ren, Z., and C. Du. 2023a. A review of machine learning state-of-charge and state-of-health estimation algorithms for lithium-ion batteries. *Energy Reports* 9:2993–3021. doi:10.1016/J.EGYR.2023.01.108.

Ren, Z., and C. Du. 2023b. A review of machine learning state-of-charge and state-of-health estimation algorithms for lithium-ion batteries. *Energy Reports* 9:2993–3021. doi:10.1016/J.EGYR.2023.01.108.

Sheikh, S. S., F. A. Shah, S. O. Athar, and H. A. Khalid. 2023. A data-driven comparative analysis of lithium-ion battery. *State of Health and Capacity Estimation* 51 (1):1–11. doi:10.1080/15325008.2022.2145389.

Singh, P., C. Fennie, and D. Reisner. 2004. Fuzzy logic modelling of state-of-charge and available capacity of nickel/metal hydride batteries. *Journal of Power Sources* 136 (2):322–33. doi:10.1016/J.JPOWSOUR.2004.03.035.
Song, Y., D. Liu, H. Liao, and Y. Peng. 2020. A hybrid statistical data-driven method for on-line joint state estimation of lithium-ion batteries. *Applied Energy* 261:114408. doi:10.1016/J.APENERGY.2019.114408.
Srivastava, S., and S. Lessmann. 2018. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. *Solar Energy* 162:232–47. doi:10.1016/J.SOLENER.2018.01.005.
Tian, J., R. Xiong, J. Lu, C. Chen, and W. Shen. 2022. Battery state-of-charge estimation amid dynamic usage with physics-informed deep learning. *Energy Storage Materials* 50:718–29. doi:10.1016/J.ENSM.2022.06.007.
Tian, J., R. Xiong, and W. Shen. 2019. A review on state of health estimation for lithium-ion batteries in photovoltaic systems. *ETransportation* 2:100028. doi:10.1016/J.ETRAN.2019.100028.
Tian, J., R. Xiong, W. Shen, and J. Lu. 2021. State-of-charge estimation of LiFePO4 batteries in electric vehicles: A deep-learning enabled approach. *Applied energy*, 291. doi:10.1016/j.apenergy.2021.116812.
Wang, W., T. Hong, X. Xu, J. Chen, Z. Liu, and N. Xu. 2019. Forecasting district-scale energy dynamics through integrating building network and long short-term memory learning algorithm. *Applied Energy* 248:217–30. doi:10.1016/J.APENERGY.2019.04.085.
Wang, Q., Z. Wang, L. Zhang, P. Liu, and L. Zhou (2022). A battery capacity estimation framework combining hybrid deep neural network and regional capacity calculation based on real-world operating data. *IEEE Transactions on Industrial Electronics*. 10.1109/TIE.2022.3229350
Wu, T., Y. Huang, Y. Xu, J. Jiang, S. Liu, and Z. Li. 2022. SOH prediction for lithium-ion battery based on improved support vector regression. *International Journal of Green Energy*. 20 (3):227–36. *2022.2027773*. doi:10.1080/15435075.2022.2027773.
Xiao, B., Y. Liu, and B. Xiao. 2019. Accurate state-of-charge estimation approach for lithium-ion batteries by gated recurrent unit with ensemble optimizer. *IEEE Access* 7:54192–202. doi:10.1109/ACCESS.2019.2913078.
Xie, Y., S. Wang, G. Zhang, Y. Fan, C. Fernandez, and F. Blaabjerg. 2023. Optimized multi-hidden layer long short-term memory modeling and suboptimal fading extended Kalman filtering strategies for the synthetic state of charge estimation of lithium-ion batteries. *Applied Energy* 336:120866. doi:10.1016/J.APENERGY.2023.120866.
Xing, Y., W. He, M. Pecht, and K. L. Tsui. 2014. State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures. *Applied Energy* 113:106–15. doi:10.1016/J.APENERGY.2013.07.008.
Xiong, R., S. Ma, H. Li, F. Sun, and J. Li. 2020. Toward a safer battery management system: a critical review on diagnosis and prognosis of battery short circuit. *IScience* 23 (4):101010. doi:10.1016/J.ISCI.2020.101010.
Xiong, Y., Y. Zhu, H. Xing, S. Lin, J. Xiao, C. Zhang, T. Yi, and Y. Fan. 2023. An improved state of charge estimation of lithium-ion battery based on a dual input model. *Energy Sources, Part A: Recovery, Utilization, & Environmental Effects* 45 (1):575–88. doi:10.1080/15567036.2023.2172479.
Yang, F., W. Li, C. Li, and Q. Miao. 2019a. State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network. *Energy* 175:66–75. doi:10.1016/J.ENERGY.2019.03.059.
Yang, F., W. Li, C. Li, and Q. Miao. 2019b. State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network. *Energy* 175:66–75. doi:10.1016/J.ENERGY.2019.03.059.
Yao, S., Y. P. Xu, and E. Ramezani. 2021. Optimal long-term prediction of Taiwan's transport energy by convolutional neural network and wildebeest herd optimizer. *Energy Reports* 7:218–27. doi:10.1016/J.EGYR.2020.12.034.
Zhang, Z., Z. Dong, H. Lin, Z. He, M. Wang, Y. He, X. Gao, and M. Gao. 2021. An improved bidirectional gated recurrent unit method for accurate state-of-charge estimation. *IEEE Access* 9:11252–63. doi:10.1109/ACCESS.2021.3049944.
Zhao, R., P. J. Kollmeyer, R. D. Lorenz, and T. M. Jahns (2017). A compact unified methodology via a recurrent neural network for accurate modeling of lithium-ion battery voltage and state-of-charge. *2017 IEEE Energy Conversion Congress and Exposition, ECCE 2017*, 2017 January, 5234–41. 10.1109/ECCE.2017.8096879
Zhou, X., J. Feng, and Y. Li. 2021. Non-intrusive load decomposition based on CNN–LSTM hybrid deep learning model. *Energy Reports* 7:5762–71. doi:10.1016/J.EGYR.2021.09.001.
Zou, Y., X. Hu, H. Ma, and S. E. Li. 2015. Combined state of charge and state of health estimation over lithium-ion battery cell cycle lifespan for electric vehicles. *Journal of Power Sources* 273:793–803. doi:10.1016/J.JPOWSOUR.2014.09.146.

Applied Chemical Engineering (2023) Volume 6 Issue 2
doi: 10.24294/ace.v6i2.2123

REVIEW ARTICLE

Advancement in steam reforming of methanol to produce hydrogen: A review

Ravindra Kumar[1], Anil Kumar[1,2,*]

[1] *Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India*

[2] *Centre for Energy and Environment, Delhi Technological University, Delhi 110042, India*

*** Corresponding author:** Anil Kumar, anilkumar76@dtu.ac.in

ABSTRACT

The main deliberation of this review paper is on metallic catalysts, including Cu-based catalysts, with distinct formulations and compositions, utilized for steam reforming of methanol (SRM). The review critically examines the performance of these catalysts, considering the active components, supports, promoters, and their interactions. Additionally, the review identifies and elucidates the various kinds of reaction mechanisms and routes involved in SRM. This comprehensive analysis provides valuable insights into the progress of well-organized and effective catalysts for SRM. To achieve high yields of H_2, it is crucial to conduct a fundamental study of the role of copper as a component in both mono and multimetallic systems, as well as the nature of support. These factors are essential to understand the catalytic mechanisms involved in the steam reforming of methanol and to develop effective strategies for optimizing hydrogen production. Therefore, a thorough investigation of copper-based catalysts and their interaction with the support material is essential for the development of highly efficient steam reforming processes.

Keywords: methanol reforming; hydrogen production; catalyst reaction scheme

ARTICLE INFO

Received: 9 May 2023
Accepted: 28 June 2023
Available online: 16 August 2023

COPYRIGHT

Copyright © 2023 by author(s).
Applied Chemical Engineering is published by EnPress Publisher, LLC. This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).
https://creativecommons.org/licenses/by-nc/4.0/

1. Introduction

In the present day, achieving stability between economic evolution, growth, and sustainability while safeguarding the environment poses a monumental challenge to the world. Energy and the environment are intricately linked, and the path toward achieving this balance grows more complex each day. According to a report in 2019 by the International Energy Agency (IEA), a 1% increase in global economic output results in a 0.5% rise in CO_2 emissions annually[1]. In recent decades, there has been a significant focus on using non-carbon renewable sources, such as solar, wind, hydrothermal, and tidal energy, as well as hydrogen energy. However, all of these sources, except hydrogen, experience availability issues throughout the year[2]. Hydrogen is becoming an increasingly popular alternative to traditional energy sources in fuel cells, thanks to its higher energy density and lack of greenhouse gas emissions[3]. Specifically, when it comes to energy sources for transportation, polymer electrolyte membrane fuel cells (PEMFC) are extensively utilized[4]. However, unlike fossil fuels that can be directly extracted from the environment, hydrogen needs to be obtained through different reforming reactions, including steam reforming, auto-thermal reforming, and partial oxidation[5]. Methane is often used as a feedstock for hydrogen generation, but this process has significant

drawbacks when it comes to industrial applications. It can be complex, and there are safety concerns related to hydrogen storage and transportation[6]. The majority of hydrogen production worldwide, particularly for the refinery and chemical industries is achieved through steam reforming or partial oxidation of fossil fuels like naphtha or natural gas. This process is well-established and widely used for various feedstocks, including LPG, methanol, diesel and jet fuel[7].

Currently, natural gas reforming is considered more cost-effective and efficient than steam reforming of methanol (SRM) for the production of H_2. However, relying on natural gas reforming is not a sustainable and environmentally friendly solution in the long term[8]. The process reforming of natural gas involves high temperatures, which results in significant CO_2 production and purification requirements for hydrogen production. This translates into higher assets funds, functional charges, and CO_2 radiations[9]. Byun et al.[10] revealed that the lowest cost for hydrogen production was 2.78 USD per kilogram of hydrogen, achieved with a reactant cost of 100 USD per ton and a cost exponent of 0.6, it remains higher compared to the cost of hydrogen produced through SMR (steam methane reforming), which is 2.30 USD per kilogram. However, the analysis of various scenarios provided valuable insights into the steps necessary to achieve economic feasibility in hydrogen production.

In modern centuries, there has been a growing interest in using methanol reforming as an attractive option for hydrogen production. Methanol has several advantages, including its small sulfur content, higher volumetric energy density, higher hydrogen/carbon ratio of 4:1, and biodegradable properties[11]. In the context of hydrogen extraction, methanol stands out as the preferred choice among other alcohols due to its ability to undergo the process at lower temperatures, typically within the range of 200 °C–300 °C. This is attributed to the absence of C–C bonds in methanol's molecular structure. In contrast, the reforming of ethanol and methane necessitates higher temperatures, around 400 °C and 500 °C, respectively[12].

The demand for methanol has been growing at a compound annual growth rate (CAGR) of 6% since 2014, mainly due to its increasing usage in various industries. Recent studies have shown that 40% of the methanol formed is used in the area of energy region[13]. The reforming technologies for hydrocarbons and alcohols have made it possible to generate hydrogen in situ, thereby overcoming the challenges of hydrogen storage and transportation[12,14]. The US Department of Energy (DOE) has established a goal to achieve hydrogen production at a charge of 4 USD/kg or lower, employing emerging technologies[1]. By 2030, it is projected that the cost of hydrogen extraction will see a further reduction to 1.8 USD/kg. Additionally, hydrogen is expected to fulfill approximately 15% of the global energy demand (GED) during that timeframe[15].

During the steam reforming of methanol (SRM) process, in which methanol is converted into hydrogen and other products, the primary product is hydrogen, along with small amounts of carbon dioxide and CO, unreacted H_2O, and CH_4. However, the generation of CO needs to be restricted, as concentrations higher than 10 ppm can damage the catalyst[16]. Therefore, the design and performance of the catalyst play a critical role. An efficient catalyst should have higher activity and selectivity to yield a higher H_2 production, thereby minimizing CO production. Additionally, the catalyst should be resistant to sintering and have the longest feasible stability[17]. Extensive investigations have been carried out on copper and palladium-based catalysts in the field of steam reforming of methanol (SRM). Among these catalysts, copper-based ones are highly favored for their remarkable activity and selectivity. However, a significant concern associated with copper-based catalysts is their susceptibility to catalytic deactivation caused by thermal sintering[18,19].

The attainment of the catalyst through the reforming reaction is dependent on both the metal component and the support material[20]. Promoters, such as acid oxides or metals of alkaline, are critical in removing unwanted reaction products. To increase the efficiency of the reaction, it is also necessary to understand catalyst deactivation caused by coking or sintering[21,22]. In a study by Song and Ozkan[23], it was proposed

that employing an improved performance reusable catalyst could offer a cost-effective solution for steam reforming of methanol (SRM). The growing interest in the SRM is apparent from the increasing number of publications on the topic each year (as shown in **Figure 1)**.

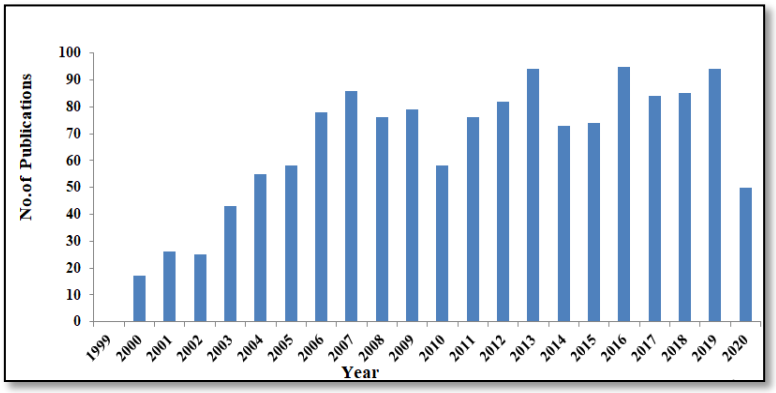


Figure 1. Publications detail on steam reforming of methanol to H_2 production from various databases, modified from Reference [1].

The cost of SRM is impacted by the process of separating hydrogen from the product mixture. Membrane reactors are more efficient than traditional reactors because they allow for the reaction and H_2 separation from CH_4, CO_2, CO, and other known by-products to occur within the same reactor, eliminating the need for additional processing steps. By bypassing the water gas shift (WGS) reaction, membrane reactors are more cost-effective. This is because they can perform both the reaction and separation simultaneously, reducing the cost and time associated with the separation process. This information is supported by research that shows the efficiency and cost-effectiveness of membrane reactors in SRM[24]. Numerous kinds of membrane reactors have been employed for SRM, including fluidized bed membrane reactors (FBMRs), packed bed membrane reactors (PBMRs), micro-porous SiO_2 membranes, and micro membrane reactors[25]. In recent years, research has also focused on disseminated feeding membranes that exhibit selectivity nearby H_2 and CO[26]. This approach has shown promise in enhancing the efficiency and selectivity of hydrogen separation in SRM. These countless varieties of membrane devices have been studied and evaluated, and their effectiveness has been documented in the literature.

Concerning its practical implementation, a study conducted by Pashchenko[27] investigated the utilization of low-grade heat for methanol decomposition. The findings demonstrated that the use of low-grade heat for methanol decomposition can lead to an efficiency improvement of up to 5% compared to utilizing the same heat for steam generation. Furthermore, the study highlighted that the efficiency of a combined cycle increases with higher gas turbine inlet temperatures.

Pashchenko et al.[28] conducted a thermodynamic analysis using the Aspen Hysys platform to evaluate the efficiency of various fuels, including methanol, ethanol, n-butanol, and glycerol, in thermochemical recuperation (TCR) systems. The analysis utilized the Gibbs free energy minimization method and investigated a wide temperature range of 400 K to 900 K, with a steam-to-fuel ratio of 1 and a pressure of 1 bar. The study focused on determining the maximum fuel conversion levels achieved. Notably, specific temperatures were found to yield the highest fuel conversion: 600 K for methanol, 730 K for ethanol, 860 K for n-butanol, and 890 K for glycerol.

Similarly, the study revealed that the efficiency of thermochemical heat recuperation systems is significantly influenced by both the composition of the initial reaction mixture and the process temperature. Notably, the maximum efficiency for steam reforming in these systems varied at different temperatures: 600 K for methanol, 700 K for ethanol, and 900 K for methane. These findings underscore the importance of

carefully considering the specific reaction mixture and process temperature to optimize the efficiency of thermochemical heat recuperation systems[29].

In the concerning of pressure drop analysis, a combination of numerical simulations and experimental investigations was employed to analyze the pressure drop and loss factor in thermochemical recuperators that incorporated catalysts of varying shapes. The reaction space within these recuperators was modeled as a porous packed bed utilizing a $NiAl_2O_3$ catalyst, and the commercial software ANSYS Fluent was utilized for the simulations. A comparison between the experimental and simulation outcomes for pressure drops revealed a strong correlation, with a discrepancy of less than 8%. Furthermore, it was observed that the pressure loss in the packed bed displayed an approximately linear relationship with the bed depth[30].

Using the RBD algorithm (rigid body dynamics), a numerical calculation was performed to investigate the flow characteristics in a realistic computational domain. The results of the analysis indicated that when the pore size is less than 0.5 mm, the flow through the particle-filled porous medium is negligible. On the other hand, when the pore size exceeds 0.5 mm, a substantial flow occurs through the porous elements of the fixed-bed reactor[31].

The main aim of this paper is to offer a thorough and inclusive examination of the recent progressions in hydrogen production utilizing the steam reforming of methanol technique. The study focuses on Cu-based catalysts and analyzes their performance along with the factors affecting their reactivity. Additionally, the literature review covers the classification of SRM, its usefulness as a hydrogen storage medium, and its potential applications.

2. Literature

2.1. Classification of steam reforming of methanol (SRM)

SRM is a chemical process used to produce hydrogen gas by reacting methanol with steam. This process is highly endothermic, meaning it requires a lot of energy to be supplied in the form of heat. SRM is often preferred over other methods of producing hydrogen such as partial oxidation of methanol (POM), oxidative steam reforming of methanol (OSRM), and sorption-enhanced steam methanol-reforming because it results in a higher yield of hydrogen. There are several categories of SRM as shown in **Figure 2**, including thermo-catalytic reforming, photo-catalytic reforming, aqueous phase reforming, and high-temperature reforming. In general, SRM is represented by the following chemical Equation (1). **Figure 2** shows the various kinds of hydrogen production methods from steam reforming of methanol.

$$\text{SRM: } CH_3OH + H_2O \rightarrow CO_2 + 3H_2;\ \Delta H^{o}_{298\,K} = +49.2\left(\frac{kJ}{mol}\right) \quad (1)$$

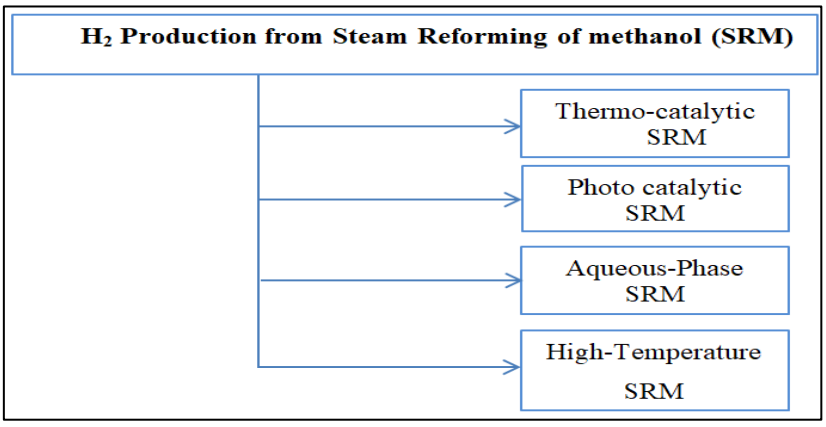


Figure 2. Classification of SRM methods.

2.1.1. Thermo-catalytic SRM

It is a commonly used method in industry for hydrogen production. However, side reactions are a common occurrence, especially when the lower ratio of steam/methanol. These side reactions often include the decomposition of methanol and the WGS reaction, leading to the production of methane, methyl formate, dimethyl ether, and other by-products[32].

$$\text{Dehydration: } 2CH_3OH \rightarrow (CH_3)_2O + H_2O; \Delta H^{o}_{298\,K} = -23\left(\frac{kJ}{mol}\right) \quad (2)$$

$$\text{Methanation: } 2CH_3OH \rightarrow CH_4 + 2H_2 + CO_2; \Delta H^{o}_{298\,K} = -66.9\left(\frac{kJ}{mol}\right) \quad (3)$$

$$\text{Dehydrogenation: } 2CH_3OH \rightarrow HCO_2CH_3 + 2H_2; \Delta H^{o}_{298\,K} = +64.4\left(\frac{kJ}{mol}\right) \quad (4)$$

$$\text{Formaldehyde synthesis: } CH_3OH \rightarrow CH_2O + H_2; \Delta H^{o}_{298\,K} = +84.7\left(\frac{kJ}{mol}\right) \quad (5)$$

$$\text{Carbonation: } CH_3OH \rightarrow C + H_2O + H_2; \Delta H^{o}_{298\,K} = +159.5\left(\frac{kJ}{mol}\right) \quad (6)$$

Wang et al.[33] were involved in a study on the thermodynamic performance of fuel cells via a poly-generation system that incorporated solar, waste heat, fuel cells, and SRM. The simulation results indicated that the system achieved an efficiency of 73.7% during the summer and 51.7% during the winter. In a separate study, Hosseini et al.[34] developed a poly-generation system that utilized a molten carbonate fuel cell (MCFC) system with an SRM, a heat and power cycle, and a methanol generation process using the distillation technique. The system demonstrated the capacity to generate electricity, warm water, and methanol concurrently, resulting in an impressive gross exergy efficiency of 83.7%. Schuller et al.[35] created a new type of heat exchanger reactor specifically for SRM. In their study, it was reported that this reactor employed waste heat from a fuel cell to drive the reforming reaction, leading to the thorough transformation of methanol at a temperature of 180 °C.

2.1.2. Photo catalytic SRM

It is a new type of technology that uses photon energy to reduce the temperature of the reaction. Sun et al.[36] studied SRM by lowering the reaction temperature with photon energy. Through a comparison of process outcomes with and without light, it was found that photon energy promotes the dissociation of formate into CO_2 and H_2. With no evidence of carbon monoxide, the rate of H_2 generation was 50.6 mmol/(g h) at 200 °C and 76.2 mmol/(g h) at 210 °C. Chiarello et al.[37] reported SRM over Pt/TiO_2 retaining in situ diminished entire reflection IR spectroscopy presented that TiO_2 physicochemical were pretentious from Pt nanoparticles.

2.1.3. Aqueous-phase SRM

It is an innovative technology that converts various compounds such as alcohol, sugar, and polyol into H_2 and CO_2 by placing them in an aqueous solution. The utilization of this process is regarded as an environmentally friendly technique for converting biomass into H_2 and CO_2 at lower reaction temperatures[38]. The key factor for the successful aqueous phase of methanol-reforming is the creation of a durable catalyst that can produce high hydrogen yields while minimizing the absorption of alkanes and CO[39]. The hydrogenation process of CO and CO_2 can lead to methane development.

$$CO + 3H_2 \rightarrow CH_4 + H_2O; \Delta H^{o}_{298\,K} = -253.70\left(\frac{kJ}{mol}\right) \quad (7)$$

$$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O; \Delta H^{o}_{298\,K} = -165\left(\frac{kJ}{mol}\right) \quad (8)$$

Davda et al.[40] utilized the catalyst (Pt/Al_2O_3) in their research on ethylene glycol and the methanol aqueous phase reforming. Their findings indicated that hydrogen had maximum activity and selectivity, with nearly 100% efficiency. Similarly, the catalyst $Ni/25\%CeO_2{-}ZrO_2$ was employed in investigating aqueous

phase SRM. The research showcased a methanol conversion rate exceeding 50% and a hydrogen efficiency surpassing 40%[41]. Cortright et al.[42] also reported on the feasibility of H_2 extraction from alcohols at approximately 227 °C via aqueous phase SRM. These studies suggest the potential for efficient hydrogen production through AP-SRM using various catalysts.

2.1.4. High-temperature SRM

This technique explores the utilization of bimetallic catalysts, specifically iron and copper supported on Al_2O_3-Zn-ZrO_2, for the extraction of hydrogen. The investigation revealed that hydrogen extraction without any detectable carbon monoxide could be achieved at a temperature of 500 °C. The catalytic activity was evaluated across a temperature range of 150 °C to 500 °C. The research team observed an intriguing synergy that led to a decrease in particle size and improved reducibility for a bimetallic catalyst with an equal Fe/Cu ratio of 1:1. Furthermore, when the copper content exceeded 75%, the selectivity towards CO was practically eliminated. By conducting DRIFTs analysis, the researchers were able to identify that the existence of Fe_2O_3 as the surface species on the catalyst played a key role in the transformation of produced CO to CO_2. The results of the analysis indicated that Fe_2O_3 acted as an active component in facilitating the chemical transformation of CO into CO_2. These findings suggest the potential for efficient H_2 production using bimetallic catalysts through iron and copper[42,43].

2.2. Chemical hydrides as methanol

Chemical hydrides and metal hydrides share the characteristic of chemically bonding hydrogen, but their properties differ significantly due to the presence of lighter elements in chemical hydrides. Significantly, chemical hydrides exhibit a liquid state under normal conditions, which offers several advantages. It simplifies the transportation and storage of these hydrides and facilitates heat and mass transfer through dehydrogenation and hydrogenation approaches. Some chemical hydrides, including formic acid, methanol, and ammonia are already commonly produced through natural gas and used for purposes beyond hydrogen storage. This pre-existing infrastructure for production, handling, and transport is advantageous. Furthermore, utilizing hydrogen obtained through water electrolysis instead of natural gas reforming for the production of these bulk chemicals can contribute to a reduction in fossil fuel consumption. While certain chemical hydrides have been proposed as substitutes for hydrogen, this article does not aim to compare these fundamentally distinct pathways[44].

Methanol (CH_3OH) exhibits hydrogen storage capacities of 12.5% (wt) and 99 kg/m^3 in terms of gravimetric and volumetric measurements, respectively, making it the simplest alcohol with the potential for hydrogen storage. One of the most widely discussed methods for producing renewable methanol is over-carbon dioxide (CO_2) hydrogenation[45]. There are several ways to release hydrogen from methanol, including through steam reforming where it reacts with water, partial oxidation where it reacts with oxygen, or via methanol thermolysis (decomposition)[46]. Among these methods, steam reforming stands out as advantageous because it yields 3 moles of H_2 per mole of CH_3OH, whereas the other two reactions result in only 2 moles of H_2 per mole of CH_3OH[46,47].

To facilitate the methanol steam reforming, external heating is required due to the endothermic nature of the reaction and the evaporation of water and methanol. Typically, a catalyst (Cu/ZnO/Al_2O_3) is employed at temperatures ranging from 230 °C–330 °C[46]. The production of methanol through the synthesis of CO_2 and hydrogen has reached a commercial stage, and the “George Olah Renewable Methanol Plant” has been operational in Iceland since 2011, making it the world’s first facility of its kind. The same type of catalyst used in conventional natural gas-based development can be used in the CO_2-based process[45]. The reaction to form methanol from CO_2 and hydrogen is exothermic, and temperatures around 220 °C–280 °C and pressures (10–80 bar) are normal reactor conditions. The methanol-water mixture produced from this process can be

used to hydrogen yield directly through the reaction of steam reforming, without the need for the energy-intensive separation of developed methanol and water by distillation[48]. In this scenario, the overall process can be streamlined, resulting in a significant surplus of heat during the synthesis of methanol.

2.3. Utilization of hydrogen production in associated sectors

Hydrogen can be viewed as a feasible substitute fuel for engines originally intended to operate on different fuels due to its broad flammability range, which allows for precise regulation of engine power[49]. Chehade et al.[50] have stated that, although the reactivity of liquids is not a concern, the high density of liquid hydrogen makes it a more desirable option compared to lightweight hydrogen gas. However, due to its compressibility in smaller reservoirs, hydrogen gas offers greater storage capacity[50]. In terms of storage efficiency, high-pressure reservoirs are preferred for storing hydrogen gas over liquid nitrogen because gas molecules can be packed into a smaller space, whereas liquid nitrogen requires refrigeration units and takes up more space in its liquid state. Fuel combustion can only occur in the vaporized or gaseous states, and H_2 can reach its state of gaseous form at a relatively lower temperature. The flashpoint, which is the temperature at which a fuel can ignite in the occurrence of an ignition source, is typically lower than its boiling point[51] However, low temperatures can reduce vaporization, which may result in the fuel flame not lasting without an ignition source. Consequently, it is anticipated that engines utilizing hydrogen as fuel would necessitate less sophisticated ignition and starting equipment in comparison to engines operating on alternative fuel types[52]. Moreover, hydrogen-powered engines can operate effectively in severe conditions. For example, a hydrogen-fueled vehicle was reported to continue functioning even after being stored in lower temperatures for multiple days, without the need for ignition[53]. Furthermore, hydrogen has a unique flammability range, which presents several opportunities for its use in turbine or combustion engines. This is due to the significant disparity between its lower flammability level (LFL) and higher flammability level (HFL)[54].

3. Cu-based catalyst

Copper-based catalysts are widely used in the steam reforming of methanol. To improve their selectivity and prevent deactivation, various methods have been proposed in the literature. Some of these methods involve investigating the configuration, properties, and preparation of the catalyst, including the accumulation of promoters and supports, while others focus on optimizing preparation methods. Furthermore, numerous scientists have underscored the significance of examining the interfacial synergy between supported metal catalysts and the interactions among their constituents. This approach is crucial for acquiring a comprehensive comprehension of the reaction mechanisms and deactivation processes[55].

3.1. Composition and active components

It is widely recognized that $Cu/ZnO/Al_2O_3$ serves as the commercial catalyst for the SRM reaction and the associated water-gas-shift (WGS) process employed in industrial hydrogen production. The catalyst ($Cu/ZnO/Al_2O_3$) is widely recognized as the commercial catalyst used in the SRM reaction and the associated water-gas-shift (WGS) process, which are integral to industrial hydrogen production.

As per prior research papers, copper is typically recognized as the primary energetic ingredient of the metal compound, implying that the catalytic characteristics are sturdily influenced by the chemical properties and structure of copper. Though, it is important to note that higher copper loading does not necessarily result in better activity and selectivity. The performance of the catalyst greatly depends on the synergy with supports and the interaction between various components. ZnO, for instance, is known to enhance the surface catalytic properties of Cu by promoting its reducibility, which plays an essential role play in the process of reforming.

Moreover, the process can be influenced by the acid/base characteristics of ZnO and its capacity to

enhance copper dispersion while inhibiting sintering. Utilizing porous alumina as catalyst support offers the advantage of generating a large surface area, resulting in enhanced copper dispersion, increased loading capacity, and reduced particle size. Furthermore, recent studies have focused extensively on investigating various dopants and support to enhance the overall performance of Cu-based catalysts, leading to promising advancements[56–58].

Fu et al.[59] conducted a study showcasing the effective adsorption of both CO and water through Cu doping of the Fe_3O_4 (111) surface, which in turn facilitated the steam reforming process. Furthermore, recent advancements highlighted the remarkable enhancement in the specific surface area achieved by the 2D Cu-Fe-Al-O nanosheet structure, leading to the creation of additional active sites for metallic Cu[60]. This innovative structure not only contributed to the stability of the catalyst but also enhanced its catalytic activity, showcasing the potential of Cu-Fe composite materials[61].

3.2. Effects of promoters on catalytic activity and selectivity

Copper, as the energetic element, can be easily influenced by various promoters to enhance its catalytic performance. The desirable objectives in producing highly active catalysts are achieving well-dispersed and small copper particles, as well as maintaining stable physicochemical properties. In addition, these promoters can also lead to improved selectivity, which is an added benefit[55]. **Table 1** presents the impact of these promoters on the performance of SRM catalysts. The impact of zirconia as a structural promoter has been studied extensively over the years. Recent research has led to a better understanding of the synergistic effects of zirconia, revealing that it can improve both the stability of the catalyst and the dispersion of the active phase. As a result, this structural promoter has been shown to significantly enhance both the activity and selectivity of SRM[55].

Azenha et al.[62] conducted a study to examine how the SRM behavior is affected by the structure of the ZrO_2 support, specifically comparing the impacts of the monoclinic and cubic forms. The utilization of ZrO_2-monoclinic as a support in CuePd/ZrO_2-monoclinic catalyst resulted in significantly improved metal phase dispersion, leading to a remarkable increase of up to 13 times in activity and a doubling of selectivity compared to the CuePd/ZrO_2-cubic catalyst, as observed in the study. By incorporating zirconia into the conventional CuO/ZnO/Al_2O_3 catalyst, the researchers observed enhanced stability, which was attributed to the presence of surface oxygen sites on zirconia clusters and an improved ability to adsorb methanol. This finding indicates that the addition of zirconia has beneficial effects on the catalyst's performance[63]. The inclusion of zirconium was found to induce microstrain in CuO and ZnO, resulting in a decrease in their crystallite size and limiting their growth. Moreover, Zr-based catalysts exhibited a higher quantity of metallic copper during the reduction process, even at lower temperatures. These observations highlight the influence of zirconium on the structural and chemical properties of the catalyst[64].

Table 1. Promoters effect on various categories of catalyst (Cu-based) for SMR.

Catalyst	T (ºC)	GHSV, LHSV or WHSV (h^{-1})	X_{MeOH} (%)	Activity (mmol/g_{cat}/h)	S_{H2} (%)	S_{CO} (%)	Reference
$Cu_{33}Fe_{19}Al_{15}O_{33}$	240	1760 (g)	94.3	225	100	-	[60]
CuPd/ZrO_2-m	260	2400 (g)[a]	>80	40	87	5	[62]
Cu/ZnO/ZrO_2	250	21,600 (g)[a]	88.6	806.41[b]	75	NA	[64]
$CuMnO_x$	400	12,000 (g)[a]	75	414	-	5	[65]
$CuZn_3GaZr$	275	2200 (g)	75	836	-	0.3	[66]
CueGa/ZnO	320	3600 (g)	96	720	74	NA	[67]
CuO/ZnO/CeO_2/Al_2O_3	260	1000 (g)[a]	98	-	65	NA	[68]
CuO/ZnO/Al_2O_3/ZrO_2/CeO_2	220	10,000 (g)[a]	100	-	75	1	[69]
CueMgO/Al_2O_3	220	2923 (g)[a]	100	-	99.4	<0.15	[70]

Table 1. (*Continued*).

Catalyst	T (ºC)	GHSV, LHSV or WHSV (h^{-1})	X_{MeOH} (%)	Activity (mmol/g_{cat}/h)	S_{H2} (%)	S_{CO} (%)	Reference
$Cu/Ni/Zr/Ce_{0.1}$	330	172 (g)	60	1000	100	NA	[71]
Cu/SiO_2	400	2400 (g)[a]	97	-	72–77	<11	[72]
$CueNi/LaZnAlO_4$	300	20.8 (w)	90	-	>80	<2	[73]
CueFe/ATP	300	3.52 (w)	100	-	98.25	2.55	[74]

The unit for "a" is mL/gcat/h (milliliters per gram of catalyst per hour).
The unit for "b" is mol/molmetal/h (moles per mole of metal per hour).

3.3. Effects of Cu-based catalyst preparation methods

Apart from utilizing various promoters and combinations of active metals, the method of preparation has also been found to significantly impact catalytic performance, a wide range of preparation approaches have been informed via literature for different types of materials. While co-precipitation, sol-gel synthesis, hydrothermal synthesis, and impregnation are commonly used methods, some authors have also proposed novel approaches or modifications to conventional methods to improve the catalytic activity and selectivity of the synthesized materials[50]. Hosseini et al.[75] systematically investigated the effects of the preparation method on the surface area and performance of their materials. They studied ZrO_2/CeO_2 doped $CuO/ZnO/Al_2O_3$ without specifying it as a catalyst.

Investigations have been conducted by researchers to examine the influence of various preparation techniques on factors such as surface area, metal dispersion, and particle size of catalysts containing copper. One method that has received attention is the urea nitrate combustion (UNC) method, which is known for its fast and easy synthesis of nano-catalysts. Several studies have shown that this method can result in catalysts with improved metal dispersion, smaller particle size, and greater surface region, which can lead to enhanced catalytic performance[76]. Yang et al.[77] examined different preparation methods for CuO/CeO_2 catalysts, emphasizing the importance of ceria morphology in influencing catalytic performance. This underscores the importance of selecting appropriate synthesis methods to optimize active site size, dispersion, and morphology to achieve desired properties.

3.4. Catalyst performance by supports

It should also be noted that the quality of the support material plays a crucial role in ensuring the effectiveness and stability of catalysts. Properly designed supports, when combined with specific modification techniques, can lead to the development of catalysts with extremely high surface area and metal dispersion[78,79].

In the study conducted by Tajrishi et al.[78], SBA-15 mesoporous substantial was produced using a hydrothermal arrangement and utilized as a support for Cu-based catalysts. The research findings highlighted that SBA-15 is a highly suitable support material for promoting the dispersion of Cu due to its exceptional stability and maximum surface region. Opting for suitable support materials with a high surface area is vital to enhance the dispersion of active metals. Support materials like MCM-41, SBA-15, carbon nanotubes (CNTs), spinel, and silica possess surface areas nearly ten times larger than those of compound catalysts. A previous study reported that an MCM-41-supported Cu-based catalyst, with a surface area of 662 m^2/g, demonstrated excellent stability and resistance to deactivation[79].

Acetic acid modification of MCM-41 before metal impregnation is effective in reducing particle size, leading to a significant improvement in dispersion and reduction behavior, as demonstrated in previous research[80].

Optimizing the catalytic performance of SRM relies on the careful selection of suitable support material with a high surface-to-volume ratio. This characteristic plays a pivotal role in facilitating the improved

dispersion of active components on the surface of the support. The uniform mesopores, high BET surface area, and exceptional thermal stability of mesoporous molecular sieves like SBA-15 and MCM-41 have contributed to their widespread acceptance[80].

Moreover, the surface area and pore volume of 3D metal foam, felt, and porous ceramics can be modified to suit various catalyst loadings. However, it is imperative to conduct additional investigations into the size and pores distribution, considering the dimensions of reactant molecules and their transport characteristics. An illustration of this is seen in the case of SBA-15, which possesses mesopores with sizes ranging from 3 to 30 nm, making it well-suited for the reforming of larger molecules such as tars. In contrast, traditional microporous catalysts are not suitable for this application[78]. Additionally, the thermal conductivity and permeability of catalysts can be influenced by gradient and cascading structures, ultimately affecting their performance. Therefore, it is important to consider the current application scenarios and not generalize that porosity and higher specific surface zone always lead to improved catalyst performance.

3.5. Catalyst deactivation

Despite exhibiting high catalytic activity, Cu-based catalysts are prone to deactivation due to various factors such as changes in oxidation state, sintering of the catalyst, or deposition of coke[57], as depicted in **Figure 3**. Furthermore, the presence of foreign species such as chloride and sulfur in the feed mixture can lead to the poisoning of the catalyst[81]. The research conducted by Fasanya et al.[82] demonstrated that the selectivity of side reactions was significantly influenced by the ratio of methanol to water. When the feed mixture lacked sufficient water, there was a higher selectivity towards CO and CH_4, and coke formation occurred at temperatures exceeding 300 °C.

Tonelli et al.[83] perceived that the deactivation of Cu-based catalysts can result from two concurrent processes. Firstly, the deposition of carbonate species, which is a reversible phenomenon. Secondly, the sintering of the active phase, which is an irreversible process that is more likely to occur at higher temperatures. The catalytic system, however, can undergo a simple regeneration process using either a 400 °C air flow or an inert flow at the reaction temperature. This regeneration procedure leads to nearly identical initial conversion rates, indicating the effective restoration of catalytic activity.

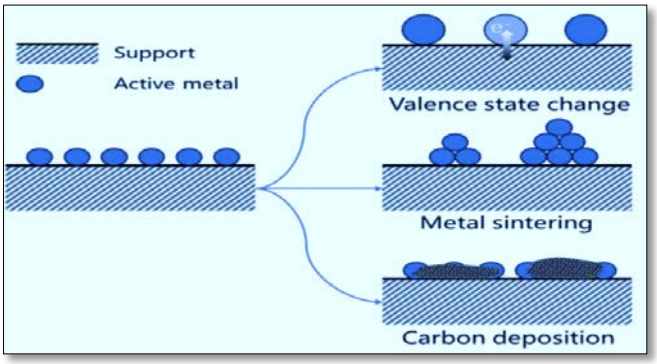


Figure 3. Various kinds of catalysts deactivation[55].

3.6. Reaction mechanism

Extensive research has been conducted on the reaction mechanism of SRM over Cu-based catalysts over a prolonged period. Initially, it was believed that the MD and WGS reactions were responsible for the reaction process[81,84].

$$CH_3OH \leftrightarrow CO + 2H_2 \quad (9)$$

$$CO + H_2O \leftrightarrow 2H_2 + CO_2 \quad (10)$$

Contrary to the initial assumption that the steam reforming of methanol (SRM) reaction is solely driven by the reaction of water gas shift and methanol decomposition, it has been discovered that the resulting product should have a higher concentration of CO, which is not observed as the produced hydrogen is nearly pure. Subsequent investigations have indicated that the steam reforming of methanol (SRM) reaction exhibits a low yield of carbon monoxide (CO) at lower temperatures, which is influenced by factors such as the steam-to-carbon ratio and catalyst properties. Further research suggests the involvement of alternative reaction pathways, including the participation of methoxy species, methyl formate, dioxymethylene, and formate intermediates[85].

Takezawa et al.[86] proposed a potential pathway for the SRM reaction mechanism involving a stepwise dehydrogenation process leading to the establishment of formaldehyde and formic acid.

$$CH_3OH \rightarrow HCHO + H_2 \quad (11)$$

$$HCHO + H_2O \rightarrow HCOOH + H_2 \quad (12)$$

$$HCOOH \rightarrow H_2 + CO_2 \quad (13)$$

The reaction rate may be restricted by the dehydrogenation step of methanol to form formaldehyde[85]. Nevertheless, there is some inconsistency in the explanations for how CO is formed. While Shishido et al.[87] supported a pathway where CO_2 is produced directly and provided evidence for a stepwise dehydrogenation pathway of methanol to form formaldehyde and formic acid, ultimately resulting in the production of H_2 and CO_2. However, there are differing opinions on the origin of CO. Some scientists argue that adsorbed formaldehyde is directly decomposed to generate H_2 and CO, which then participates in the water-gas shift reaction[85].

$$HCHO \rightarrow CO + H_2 \quad (14)$$

It is worth noting that while decomposition occurs at a slower rate compared to reforming, it still needs to be considered in most kinetic models. The endothermic nature of the decomposition reaction has a more significant impact in the presence of water, especially at higher temperatures as predicted by thermodynamics. Another proposed mechanism involves a complex process that includes a methyl formate intermediate[81].

$$2CH_3OH \rightarrow HCOOCH_3 + 2H_2 \quad (15)$$

$$HCOOCH_3 + H_2O \rightarrow CH_3OH + HCOOH \quad (16)$$

The Langmuir-Hinshelwood mechanism proposes that the initial step in methanol conversion involves the breaking of the O–H bond, leading to the formation of surface methoxy species. Subsequently, the methoxy species undergo dehydrogenation to generate formaldehyde[85].

In their work, Frank et al.[88] presented a comprehensive catalytic cycle that encompasses two types of active sites, one responsible for hydrogen adsorption and the other for the adsorption of all other intermediates. Hammoud et al.[89] illustrated a potential pathway, shown in **Figure 4**, for the production of methyl formate. In this process, formaldehyde, produced through dehydrogenation, reacts with a methoxy group. This interaction gives rise to an intermediate, which subsequently decomposes with the involvement of hydroxyl groups, resulting in the formation of methoxy and formate groups. Another pathway involves the dimerization of methoxy groups, leading to the synthesis of methyl formate.

Despite numerous studies, a clear consensus on the kinetics and reaction mechanism of SRM has yet to be reached. However, significant progress has been made in utilizing these studies to develop specialized catalyst formulations. In recent years, the pathway through a methyl formate intermediate has gained widespread acceptance as the most feasible mechanism for SRM.

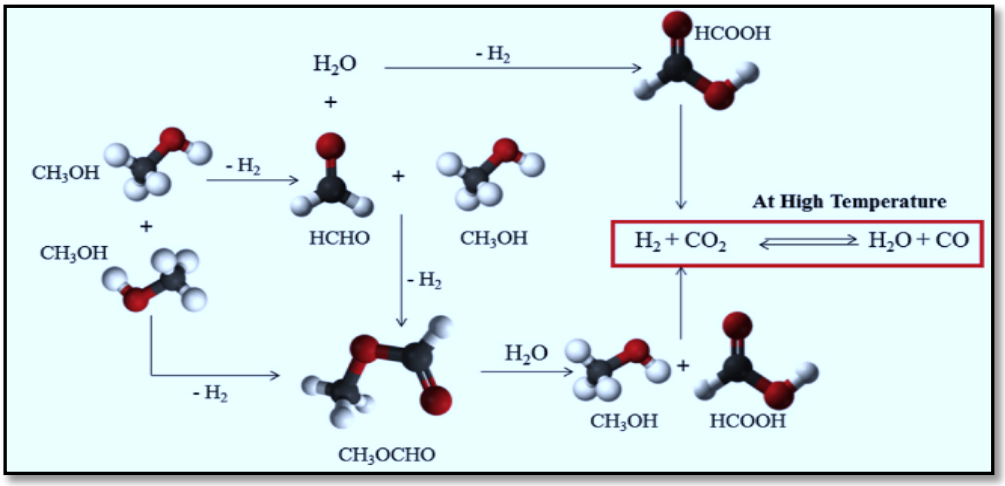


Figure 4. Reaction pathways of steam reforming of methanol[55,89].

4. Future research

The primary hurdle encountered in the SRM process is the deactivation of the catalyst caused by coking or metal sintering. To address this concern, it is crucial to develop and fabricate a catalyst that exhibits, reduced coke formation, enhanced catalytic activity, smaller particle size prolonged stability even distribution of metal particles. These catalyst properties are essential for achieving optimal hydrogen yield at lower temperatures. This requires advanced research focused on developing such catalysts using various promoters and supports. Additionally, reducing the reforming temperature and producing pure hydrogen can make the process more cost-efficient. Advanced materials like CNT (Carbon nanotubes) and SAA (single-atom alloys) can be explored to achieve these objectives, along with amalgamating various reforming reactions to enhance economic viability, for example, by utilizing heat produced through an exothermic reaction for SRM. To achieve high yields of hydrogen, it is of utmost importance to gain a deep understanding of the fundamental role played by copper in both mono and multimetallic systems, as well as the characteristics of the supporting material. The interactions between the metal and the support, specifically the transfer of oxygen from reducible oxides to the metal sites, need to be thoroughly investigated to assess their impact on catalyst performance, selectivity, and stability. Moreover, advanced characterization techniques should be employed to gain insights into the effects of alloying on selectivity in multimetallic catalysts utilizing mixed metal oxides.

5. Conclusion

The diminishing availability of fossil fuels and the growing environmental apprehensions underscore the urgent demand for clean and sustainable alternative energy sources. Hydrogen, being the most ample component on earth, holds great promise as an energy solution due to its superior energy density compared to fossil fuels on a mass basis. However, addressing the cost of production and transportation, improving energy conversion efficiency, ensuring safety, and managing the transition period for energy infrastructure requires in-depth research to develop and examine a comprehensive hydrogen supply-demand network that simultaneously meets economic, efficient, and feasible standards.

The present study undertook a comprehensive review of the literature on various approaches for steam methanol reforming (SRM) in hydrogen production. Based on the analysis conducted, the following conclusions can be drawn from this research investigation.

- Hydrogen production has traditionally depended on conventional sources such as coal, natural gas, and naphtha using steam reforming processes. However, the sustainability of these sources presents a challenge in the present era.
- In upcoming years, hydrogen is projected to have a substantial impact on carbon-neutral and carbon-negative economies. This has spurred extensive research into developing new and environmentally friendly technologies for H_2 production.
- Steam reforming methanol (SRM) has emerged as a highly promising technology that has gained significant attention. This review offers an in-depth analysis of different catalysts, their supports and promoters, the interactions among them, and the effects of various operating conditions on the performance of SRM.
- The analysis of SRM catalysis utilizing different Cu-based catalysts, as well as alternative catalysts, revealed their susceptibility to deactivation caused by coking and sintering.
- Among the Cu catalysts, catalysts (Cu/ZnO) have received significant attention, with the formation of Cu-Zn alloy considered critical for high activity. To enrich the catalytic activity and stability, different kinds of promoters such as ZrO_2, Al_2O_3 and CeO_2 have been utilized.
- The existing literature addresses different types of methanol-reforming reactions, focusing on their application in specific mobile devices powered by internal combustion engines (ICEs) with in-cylinder thermochemical fuel reforming (TFR) processes, solid oxide fuel cells (SOFCs), and polymer electrolyte membrane fuel cells (PEMFCs). The discussions consider the varied temperature requirements and inlet gas compositions associated with each technology.
- The analysis utilized the Gibbs free energy minimization method and investigated a wide temperature range of 400 to 900 K, with a steam-to-fuel ratio of 1 and a pressure of 1 bar. The study focused on determining the maximum fuel conversion levels achieved. Notably, specific temperatures were found to yield the highest fuel conversion: 600 K for methanol, 730 K for ethanol, 860 K for n-butanol, and 890 K for glycerol.
- Future investigations should prioritize achieving a consistent system output and instantaneous hydrogen yield to enhance hydrogen production and utilization.
- It is worth noting that methanol benefits in the reforming process, production, and storage making it a crucial component of the entire hydrogen network.

Author contributions

Conceptualization, RK and AK; methodology, RK; software, RK; validation, RK and AK; formal analysis, RK; investigation, RK; resources, AK; data curation, RK; writing—original draft preparation, RK; writing—review and editing, AK; visualization, AK; supervision, AK.

Acknowledgments

Authors are highly grateful to the Department of Mechanical Engineering, Centre for Energy and Environment, Delhi Technological University Delhi (India) for providing basic facilities to compile this work.

Conflict of interest

The authors declare no conflict of interest.

References

1. Ranjekar AM, Yadav GD. Steam reforming of methanol for hydrogen production: A critical analysis of catalysis, processes, and scope. *Industrial & Engineering Chemistry Research* 2021; 60(1): 89–113. doi: 10.1021/acs.iecr.0c05041
2. Olah GA. Towards oil independence through renewable methanol chemistry. *Angewandte Chemie International Edition* 2013; 52(1): 104–107. doi: 10.1002/anie.201204995
3. Wang J, Wang H, Hu P. Theoretical insight into methanol steam reforming on indium oxide with different coordination environments. *Science China Chemistry* 2018; 61: 336–343. doi: 10.1007/s11426-017-9139-x
4. Cacciola G, Antonucci V, Freni S. Technology up date and new strategies on fuel cells. *Journal of power sources* 2001; 100(1–2): 67–79. doi: 10.1016/S0378-7753(01)00884-9
5. Xu X, Liu X, Xu B. A survey of nickel-based catalysts and monolithic reformers of the onboard fuel reforming system for fuel cell APU applications. *International Journal of Energy Research* 2016; 40(9): 1157–1177. doi: 10.1002/er.3509
6. Kaftan A, Kusche M, Laurin M, et al. KOH-promoted Pt/Al_2O_3 catalysts for water gas shift and methanol steam reforming: An operando DRIFTS-MS study. *Applied Catalysis B: Environmental* 2017; 201: 169–181. doi: 10.1016/j.apcatb.2016.08.016
7. García L. Hydrogen production by steam reforming of natural gas and other nonrenewable feedstocks. In: Subramani V, Basile A, Veziroğlu TN (editors). *Compendium of Hydrogen Energy*. Woodhead Publishing; 2015. pp. 83–107.
8. Sun P, Young B, Elgowainy A, et al. Criteria air pollutants and greenhouse gas emissions from hydrogen production in US steam methane reforming facilities. *Environmental science & technology* 2019; 53(12): 7103–7113. doi: 10.1021/acs.est.8b06197
9. Cormos AM, Szima S, Fogarasi S, Cormos CC. Economic assessments of hydrogen production processes based on natural gas reforming with carbon capture. *Chemical Engineering Transactions* 2018; 70: 1231–1236. doi: 10.3303/CET1870206
10. Byun M, Lee B, Lee H, et al. Techno-economic and environmental assessment of methanol steam reforming for H_2 production at various scales. *International Journal of Hydrogen Energy* 2020; 45(46): 24146–24158. doi: 10.1016/j.ijhydene.2020.06.097
11. Ribeirinha P, Mateos-Pedrero C, Boaventura M, et al. $CuO/ZnO/Ga_2O_3$ catalyst for low temperature MSR reaction: Synthesis, characterization and kinetic model. *Applied Catalysis B: Environmental* 2018; 221: 371–379. doi: 10.1016/j.apcatb.2017.09.040
12. Basile A, Parmaliana A, Tosti S, et al. Hydrogen production by methanol steam reforming carried out in membrane reactor on Cu/Zn/Mg-based catalyst. *Catalysis Today* 2008; 137(1): 17–22. doi: 10.1016/j.cattod.2008.03.015
13. Ott J, Gronemann V, Pontzen F, et al. Methanol. In: *Ullmann's Encyclopedia of Industrial Chemistry*. American Cancer Society; 2012. doi: 10.1002/14356007.a16_465.pub3
14. Ahmed S, Krumpelt M. Hydrogen from hydrocarbon fuels for fuel cells. *International Journal of Hydrogen Energy* 2001; 26(4); 291–301. doi: 10.1016/S0360-3199(00)00097-5
15. Staffell I, Scamman D, Abad AV, et al. The role of hydrogen and fuel cells in the global energy system. *Energy & Environmental Science* 2019; 12(2): 463–491. doi: 10.1039/c8ee01157e
16. Dhar HP, Christner LG, Kush AK. Nature of CO adsorption during H_2 oxidation in relation to modeling for CO poisoning of a fuel cell anode. *Journal of the Electrochemical Society* 1987; 134(12): 3021. doi: 10.1149/1.2100333
17. Kubacka A, Fernández-García M, Martínez-Arias A. Catalytic hydrogen production through WGS or steam reforming of alcohols over Cu, Ni and Co catalysts. *Applied Catalysis A: General* 2016; 518: 2–17. doi: 10.1016/j.apcata.2016.01.027
18. Karim AM, Conant T, Datye AK. Controlling ZnO morphology for improved methanol steam reforming reactivity. *Physical Chemistry Chemical Physics* 2008; 10(36): 5584–5590. doi: 10.1039/b800009c
19. Papavasiliou J, Avgouropoulos G, Ioannides T. Steady-state isotopic transient kinetic analysis of steam reforming of methanol over Cu-based catalysts. *Applied Catalysis B: Environmental* 2009; 88(3–4): 490–496. doi: 10.1016/j.apcatb.2008.10.018
20. Kubacka A, Martínez-Arias A, Fernández-García M. Role of the interface in base-metal ceria-based catalysts for hydrogen purification and production processes. *ChemCatChem* 2015; 7(22): 3614–3624. doi: 10.1002/cctc.201500593
21. Hardiman KM, Ying TT, Adesina AA, et al. Performance of a Co-Ni catalyst for propane reforming under low steam-to-carbon ratios. *Chemical Engineering Journal* 2004; 102(2): 119–130. doi: 10.1016/j.cej.2004.03.005
22. Alenazey F, Cooper CG, Dave CB, et al. Coke removal from deactivated Co-Ni steam reforming catalyst using different gasifying agents: An analysis of the gas-solid reaction kinetics. *Catalysis Communications* 2009; 10(4): 406–411. doi: 10.1016/j.catcom.2008.10.010

23. Song H, Ozkan US. Economic analysis of hydrogen production through a bio-ethanol steam reforming process: Sensitivity analyses and cost estimations. *International Journal of Hydrogen Energy* 2010; 35(1); 127–134. doi: 10.1016/j.ijhydene.2009.10.043
24. Zhang G, Jin W, Xu N. Design and fabrication of ceramic catalytic membrane reactors for green chemical engineering applications. *Engineering* 2018; 4(6): 848–860. doi: 10.1016/j.eng.2017.05.001
25. Gallucci F, Fernandez E, Corengia P, van Sint Annaland M. Recent advances on membranes and membrane reactors for hydrogen production. *Chemical Engineering Science* 2013; 92: 40–66. doi: 10.1016/j.ces.2013.01.008
26. Liuzzi D, Fernandez E, Perez S, et al. Advances in membranes and membrane reactors for the Fischer-Tropsch synthesis process for biofuel production. *Reviews in Chemical Engineering* 2022; 38(1): 55–76. doi: 10.1515/revce-2019-0067
27. Pashchenko D. Low-grade heat utilization in the methanol-fired gas turbines through a thermochemical fuel transformation. *Thermal Science and Engineering Progress* 2022; 36: 101537. doi: 10.1016/j.tsep.2022.101537
28. Pashchenko D, Gnutikova M, Karpilov I. Comparison study of thermochemical waste-heat recuperation by steam reforming of liquid biofuels. *International Journal of Hydrogen Energy* 2020; 45(7): 4174–4181. doi: 10.1016/j.ijhydene.2019.11.202
29. Pashchenko D. Thermochemical waste-heat recuperation as on-board hydrogen production technology. *International Journal of Hydrogen Energy* 2021; 46(57): 28961–28968. doi: 10.1016/j.ijhydene.2020.11.108
30. Pashchenko D. Pressure drop in the thermochemical recuperators filled with the catalysts of various shapes: A combined experimental and numerical investigation. *Energy* 2019; 166: 462–470. doi: 10.1016/j.energy.2018.10.084
31. Pashchenko D, Karpilov I, Mustafin R. Numerical calculation with experimental validation of pressure drop in a fixed-bed reactor filled with the porous elements. *AIChE Journal* 2020; 66(5): e16937. doi: 10.1002/aic.16937
32. Thattarathody R, Artoul M, Digilov RM, Sheintuch M. Pressure, diffusion, and S/M ratio effects in methanol steam reforming kinetics. *Industrial & Engineering Chemistry Research* 2018; 57(9): 3175–3186. doi: 10.1021/acs.iecr.7b05033
33. Wang J, Wu J, Xu Z, Li M. Thermodynamic performance analysis of a fuel cell trigeneration system integrated with solar-assisted methanol reforming. *Energy Conversion and Management* 2017; 150: 81–89. doi: 10.1016/j.enconman.2017.08.012
34. Hosseini SS, Mehrpooya M, Alsagri AS, Alrobaian AA. Introducing, evaluation and exergetic performance assessment of a novel hybrid system composed of MCFC, methanol synthesis process, and a combined power cycle. *Energy Conversion and Management* 2019; 197: 111878. doi: 10.1016/j.enconman.2019.111878
35. Schuller G, Vázquez FV, Waiblinger W, et al. Heat and fuel coupled operation of a high temperature polymer electrolyte fuel cell with a heat exchanger methanol steam reformer. *Journal of Power Sources* 2017; 347: 47–56. doi: 10.1016/j.jpowsour.2017.02.021
36. Sun Z, Fang S, Lin Y, Hu YH. Photo-assisted methanol steam reforming on solid solution of Cu-Zn-Ti oxide. *Chemical Engineering Journal* 2019; 375: 121909. doi: 10.1016/j.cej.2019.121909
37. Chiarello GL, Ferri D, Selli E. In situ attenuated total reflection infrared spectroscopy study of the photocatalytic steam reforming of methanol on Pt/TiO_2. *Applied Surface Science* 2018; 450: 146–154. doi: 10.1016/j.apsusc.2018.04.167
38. Fasolini A, Cucciniello R, Paone E, et al. A short overview on the hydrogen production via aqueous phase reforming (APR) of cellulose, C6-C5 sugars and polyols. *Catalysts* 2019; 9(11): 917. doi: 10.3390/catal9110917
39. Huber GW, Dumesic JA. An overview of aqueous-phase catalytic processes for production of hydrogen and alkanes in a biorefinery. *Catalysis Today* 2006; 111(1–2): 119–132. doi: 10.1016/j.cattod.2005.10.010
40. Davda RR, Shabaker JW, Huber GW, et al. A review of catalytic issues and process conditions for renewable hydrogen and alkanes by aqueous-phase reforming of oxygenated hydrocarbons over supported metal catalysts. *Applied Catalysis B: Environmental* 2005; 56(1–2): 171–186. doi: 10.1016/j.apcatb.2004.04.027
41. Stekrova M, Rinta-Paavola A, Karinen R. Hydrogen production via aqueous-phase reforming of methanol over nickel modified Ce, Zr and La oxide supports. *Catalysis Today* 2018; 304: 143–152. doi: 10.1016/j.cattod.2017.08.030
42. Cortright RD, Davda RR, Dumesic JA. Hydrogen from catalytic reforming of biomass-derived hydrocarbons in liquid water. *Nature* 2002; 418(6901): 964–967. doi: 10.1038/nature01009
43. Sharma R, Kumar A, Upadhyay RK. Bimetallic Fe-promoted catalyst for CO-free hydrogen production in high-temperature-methanol steam reforming. *ChemCatChem* 2019; 11(18): 4568–4580. doi: 10.1002/cctc.201901062
44. Andersson J, Grönkvist S. Large-scale storage of hydrogen. *International Journal of Hydrogen Energy* 2019; 44(23): 11901–11919. doi: 10.1016/j.ijhydene.2019.03.063
45. Goeppert A, Czaun M, Jones JP, et al. Recycling of carbon dioxide to methanol and derived products-closing the loop. *Chemical Society Reviews* 2014; 43(23): 7995–8048. doi: 10.1039/c4cs00122b
46. Guczi L, Erdôhelyi A (editors). *Catalysis for Alternative Energy Generation*. Springer; 2012.
47. Richards N, Needels J, Erickson P. Autothermal-reformation enhancement using a stratified-catalyst technique. *International Journal of Hydrogen Energy* 2017; 42(41): 25914–25923. doi: 10.1016/j.ijhydene.2017.08.050

48. Klankermayer J, Wesselbaum S, Beydoun K, Leitner W. Selective catalytic synthesis using the combination of carbon dioxide and hydrogen: Catalytic chess at the interface of energy and chemistry. *Angewandte Chemie International Edition* 2016; 55(26): 7296–7343. doi: 10.1002/anie.201507458
49. Emami SD, Kasmani RM, Hamid MD, et al. Kinetic and dynamic analysis of hydrogen-enrichment mixtures in combustor systems—A review paper. *Renewable and Sustainable Energy Reviews* 2016; 62: 1072–1082. doi: 10.1016/j.rser.2016.05.029
50. Chehade G, Lytle S, Ishaq H, et al. Hydrogen production by microwave based plasma dissociation of water. *Fuel* 2020; 264: 116831. doi: 10.1016/j.fuel.2019.116831
51. Niermann M, Beckendorff A, Kaltschmitt M, Bonhoff K. Liquid organic hydrogen carrier (LOHC)—Assessment based on chemical and economic properties. *International Journal of Hydrogen Energy* 2019; 44: 6631–6654. doi: 10.1016/j.ijhydene.2019.01.199
52. Dimitriou P, Tsujimura T. A review of hydrogen as a compression ignition engine fuel. *International Journal of Hydrogen Energy* 2017; 42(38): 24470–24486. doi: 10.1016/j.ijhydene.2017.07.232
53. Yan F, Xu L, Wang Y. Application of hydrogen enriched natural gas in spark ignition IC engines: From fundamental fuel properties to engine performances and emissions. *Renewable and Sustainable Energy Reviews* 2018; 82: 1457–1488. doi: 10.1016/j.rser.2017.05.227
54. Mazloomi K, Gomes C. Hydrogen as an energy carrier: Prospects and challenges. *Renewable and Sustainable Energy Reviews* 2012; 16(5): 3024–3033. doi: 10.1016/j.rser.2012.02.028
55. Li H, Ma C, Zou X, et al. On-board methanol catalytic reforming for hydrogen production—A review. *International Journal of Hydrogen Energy* 2021; 46(43): 22303–22327. doi: 10.1016/j.ijhydene.2021.04.062
56. Yong ST, Ooi CW, Chai SP, Wu XS. Review of methanol reforming-Cu-based catalysts, surface reaction mechanisms, and reaction schemes. *International Journal of Hydrogen Energy* 2013; 38(22): 9541–9552. doi: 10.1016/j.ijhydene.2013.03.023
57. Kubacka A, Fernández-García M, Martínez-Arias A. Catalytic hydrogen production through WGS or steam reforming of alcohols over Cu, Ni and Co catalysts. *Applied Catalysis A: General* 2016; 518: 2–17. doi: 10.1016/j.apcata.2016.01.027
58. Li J, Mei X, Zhang L, et al. A comparative study of catalytic behaviors of Mn, Fe, Co, Ni, Cu and Zn-based catalysts in steam reforming of methanol, acetic acid and acetone. *International Journal of Hydrogen Energy* 2020; 45(6): 3815–3832. doi: 10.1016/j.ijhydene.2019.03.269
59. Fu Z, Wang J, Zhang N, et al. Effect of Cu doping on the catalytic activity of Fe_3O_4 in water-gas shift reactions. *International Journal of Hydrogen Energy* 2015; 40(5): 2193–2198. doi: 10.1016/j.ijhydene.2014.12.063
60. Wang L, Zhang F, Miao D, et al. Preparation of Cu-Fe-Al-O nanosheets and their catalytic application in methanol steam reforming for hydrogen production. *Materials Research Express* 2017; 4(3): 035005. doi: 10.1088/2053-1591/4/3/035005
61. Zhang F, Guo H, Wang L, et al. Porous $Al_{63}Cu_{25}Fe_{12}$ quasicrystals covered with $(Al_{11.5}Fe_{13.9}Cu_{19.7})O_{54.9}$ nanosheets. *Materials Characterization* 2019; 147: 165–172. doi: 10.1016/j.matchar.2018.10.026
62. Azenha CSR, Mateos-Pedrero C, Queirós S, et al. Innovative ZrO_2-supported CuPd catalysts for the selective production of hydrogen from methanol steam reforming. *Applied Catalysis B: Environmental* 2017; 203: 400–407. doi: 10.1016/j.apcatb.2016.10.041
63. Ahmadi F, Haghighi M, Ajamein H. Sonochemically coprecipitation synthesis of $CuO/ZnO/ZrO_2/Al_2O_3$ nanocatalyst for fuel cell grade hydrogen production via steam methanol reforming. *Journal of Molecular Catalysis A: Chemical* 2016; 421: 196–208. doi: 10.1016/j.molcata.2016.05.027
64. Sanches SG, Flores JH, da Silva MIP. Cu/ZnO and $Cu/ZnO/ZrO_2$ catalysts used for methanol steam reforming. *Molecular Catalysis* 2018; 454: 55–62. doi: 10.1016/j.mcat.2018.05.012
65. Gac W, Słowik G, Zawadzki W. Structural and surface changes of copper modified manganese oxides. *Applied Surface Science* 2016; 370: 536–544. doi: 10.1016/j.apsusc.2016.02.136
66. Liu X, Toyir J, de la Piscina PR, Homs N. Hydrogen production from methanol steam reforming over Al_2O_3- and ZrO_2-modified $CuOZnOGa_2O_3$ catalysts. *International Journal of Hydrogen Energy* 2017; 42(19): 13704–13711. doi: 10.1016/j.ijhydene.2016.12.133
67. Toyir J, de la Piscina PR, Homs N. Ga-promoted copper-based catalysts highly selective for methanol steam reforming to hydrogen; relation with the hydrogenation of CO_2 to methanol. *International Journal of Hydrogen Energy* 2015; 40(34): 11261–11266. doi: 10.1016/j.ijhydene.2015.04.039
68. Shokrani R, Haghighi M, Jodeiri N, et al. Fuel cell grade hydrogen production via methanol steam reforming over $CuO/ZnO/Al_2O_3$ nanocatalyst with various oxide ratios synthesized via urea-nitrates combustion method. *International Journal of Hydrogen Energy* 2014; 39(25): 13141–13155. doi: 10.1016/j.ijhydene.2014.06.048
69. Bagherzadeh SB, Haghighi M, Rahemi N. Novel oxalate gel coprecipitation synthesis of ZrO_2-CeO_2-promoted $CuO-ZnO-Al_2O_3$ nanocatalyst for fuel cell-grade hydrogen production from methanol: Influence of ceria-zirconia loading. *Energy Conversion and Management* 2017; 134: 88–102. doi: 10.1016/j.enconman.2016.12.005
70. Li H, Tian H, Chen S, et al. Sorption enhanced steam reforming of methanol for high-purity hydrogen production over $Cu-MgO/Al_2O_3$ bifunctional catalysts. *Applied Catalysis B: Environmental* 2020; 276: 119052. doi: 10.1016/j.apcatb.2020.119052

71. Lytkina AA, Orekhova NV, Ermilova MM, Yaroslavtsev AB. The influence of the support composition and structure ($M_XZr_{1-X}O_{2-\delta}$) of bimetallic catalysts on the activity in methanol steam reforming. *International Journal of Hydrogen Energy* 2018; 43(1): 198–207. doi: 10.1016/j.ijhydene.2017.10.182
72. Sarafraz MM, Safaei MR, Goodarzi M, Arjomandi M. Reforming of methanol with steam in a micro-reactor with Cu–SiO_2 porous catalyst. *International Journal of Hydrogen Energy* 2019; 44(36): 19628–19639. doi: 10.1016/j.ijhydene.2019.05.215
73. Khani Y, Bahadoran F, Soltanali S, Ahari JS. Hydrogen production by methanol steam reforming on a cordierite monolith reactor coated with Cu–Ni/$LaZnAlO_4$ and Cu–Ni/γ-Al_2O_3 catalysts. *Research on Chemical Intermediates* 2018; 44: 925–942. doi: 10.1007/s11164-017-3144-8
74. Cao L, Lu M, Li G, Zhang S. Hydrogen production from methanol steam reforming catalyzed by Fe modified Cu supported on attapulgite clay. *Reaction Kinetics, Mechanisms and Catalysis* 2019; 126: 137–152. doi: 10.1007/s11144-018-1493-y
75. Hosseini T, Haghighi M, Ajamein H. Fuel cell-grade hydrogen production from methanol over sonochemical coprecipitated copper based nanocatalyst: Influence of irradiation power and time on catalytic properties and performance. *Energy Conversion and Management* 2016; 126: 595–607. doi: 10.1016/j.enconman.2016.07.056
76. Talkhoncheh SK, Haghighi M, Minaei S, et al. Synthesis of CuO/ZnO/Al_2O_3/ZrO_2/CeO_2 nanocatalysts via homogeneous precipitation and combustion methods used in methanol steam reforming for fuel cell grade hydrogen production. *RSC Advances* 2016; 6(62): 57199–57209. doi: 10.1039/C6RA03858A
77. Yang S, Zhou F, Liu Y, et al. Morphology effect of ceria on the performance of CuO/CeO_2 catalysts for hydrogen production by methanol steam reforming. *International Journal of Hydrogen Energy* 2019; 44(14): 7252–7261. doi: 10.1016/j.ijhydene.2019.01.254
78. Tajrishi OZ, Taghizadeh M, Kiadehi AD. Methanol steam reforming in a microchannel reactor by Zn-, Ce- and Zr-modified mesoporous Cu/SBA-15 nanocatalyst. *International Journal of Hydrogen Energy* 2018; 43(31): 14103–14120. doi: 10.1016/j.ijhydene.2018.06.035
79. Deshmane VG, Abrokwah RY, Kuila D. Synthesis of stable Cu-MCM-41 nanocatalysts for H_2 production with high selectivity via steam reforming of methanol. *International Journal of Hydrogen Energy* 2015; 40(33): 10439–10452. doi: 10.1016/j.ijhydene.2015.06.084
80. Mohtashami Y, Taghizadeh M. Performance of the ZrO_2 promoted Cu–ZnO catalyst supported on acetic acid-treated MCM-41 in methanol steam reforming. *International Journal of Hydrogen Energy* 2019; 44(12): 5725–5738. doi: 10.1016/j.ijhydene.2019.01.029
81. Sá S, Silva H, Brandão L, et al. Catalysts for methanol steam reforming—A review. *Applied Catalysis B: Environmental* 2010; 99(1–2): 43–57. doi: 10.1016/j.apcatb.2010.06.015
82. Fasanya OO, Al-Hajri R, Ahmed OU, et al. Copper zinc oxide nanocatalysts grown on cordierite substrate for hydrogen production using methanol steam reforming. *International Journal of Hydrogen Energy* 2019; 44(41): 22936–22946. doi: 10.1016/j.ijhydene.2019.06.185
83. Tonelli F, Gorriz O, Tarditi A, et al. Activity and stability of a CuO/CeO_2 catalyst for methanol steam reforming. *International Journal of Hydrogen Energy* 2015; 40(39): 13379–13387. doi: 10.1016/j.ijhydene.2015.08.046
84. Palo DR, Dagle RA, Holladay JD. Methanol steam reforming for hydrogen production. *Chemical Reviews* 2007; 107(10): 3992–4021. doi: 10.1021/cr050198b
85. Kapran AY, Orlyk SM. Hydrogen production in methanol reforming on modified copper-zinc catalysts: A review. *Theoretical and Experimental Chemistry* 2017; 53: 1–16. doi: 10.1007/s11237-017-9495-9
86. Takezawa N, Iwasa N. Steam reforming and dehydrogenation of methanol: Difference in the catalytic functions of copper and group VIII metals. *Catalysis Today* 1997; 36(1): 45–56. doi: 10.1016/S0920-5861(96)00195-2
87. Shishido T, Yamamoto Y, Morioka H, Takehira K. Production of hydrogen from methanol over Cu/ZnO and Cu/ZnO/Al_2O_3 catalysts prepared by homogeneous precipitation: Steam reforming and oxidative steam reforming. *Journal of Molecular Catalysis A: Chemical* 2007; 268(1–2): 185–194. doi: 10.1016/j.molcata.2006.12.018
88. Frank B, Jentoft FC, Soerijanto H, et al. Steam reforming of methanol over copper-containing catalysts: Influence of support material on microkinetics. *Journal of Catalysis* 2007; 246(1): 177–192. doi: 10.1016/j.jcat.2006.11.031
89. Hammoud D, Gennequin C, Aboukaïs A, Abi Aad E. Steam reforming of methanol over x% Cu/Zn–Al 400 500 based catalysts for production of hydrogen: Preparation by adopting memory effect of hydrotalcite and behavior evaluation. *International Journal of Hydrogen Energy* 2015; 40(2): 1283–1297. doi: 10.1016/j.ijhydene.2014.09.080

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 27 2023

Aerodynamic study of various tall building under wind load

FREE

Payal Devi; Abhishek Prakash Paswan; Hemant Gautam; Rahul Kumar Meena ✉; Ritu Raj

Check for updates

AIP Conference Proceedings 2721, 060002 (2023)
https://doi.org/10.1063/5.0154002

View Online
Export Citation
CrossMark

10 August 2023 06:50:22

500 kHz or 8.5 GHz?
And all the ranges in between.
Lock-in Amplifiers for your periodic signal measurements

Find out more
Zurich Instruments

AIP Publishing

Aerodynamic Study of Various Tall Building Under Wind Load

Payal Devi[1], Abhishek Prakash Paswan[1], Hemant Gautam[1], Rahul Kumar Meena[1,2,a] and Ritu Raj[1]

[1]Department of Civil Engineering, Delhi Technological University, New Delhi, 110042, India
[2] Department of Civil Engineering, Punjab Engineering College, Chandigah,160012, India

a) Corresponding Author: rahulmeena@pec.edu.in

Abstract: High-rise structures have always fascinated people since the dawn of civilization, and they are unique in many ways, including the consideration of lateral deflections caused by lateral loads. In this paper, an attempt is made to present the results of CFD analysis on 1:100 length scale models of rectangular-shaped buildings with different side ratios, having the same height and same plan area. The study was carried out using the ANSYS CFX package for CFD simulation. Pressure coefficients on each face for different angles of attack are computed. Wind pressure variations on different faces of buildings are presented using contour plots. For the purpose of validation, the result of a regular square-plan shaped building is compared with previously experimental data for similar wind flow conditions. It has been concluded that with a change in side ratio, there is an almost negligible effect on pressure distribution and pressure coefficient on windward walls. However, there is a considerable change in the pressure coefficient on leeward as well as sideward walls. The study also revealed how a change in the angle of attack changes the magnitude of pressure coefficients.

Keywords: Tall building; pressure coefficients; numerical simulation; side ratio; vertical center line.

10 August 2023 06:50:22

INTRODUCTION

With the advancement of technology and enormous growth in the population, the need for and design of tall structures with different configurations has been a growing trend. High-rise structures have always fascinated people since the dawn of civilization, and they are unique in many ways, including the consideration of lateral deflections. Wind is a complicated phenomenon in which the motion of an individual particle is so unpredictable that one needs to be concerned about the statistical distribution of velocity rather than just simple averages. The total wind force is equal to the sum of windward pressure and leeward suction, although each has its own local effect. The designers consult appropriate wind load standards such as (AS/NZS 1170.2:2011), (ASCE 7-10), (BS 6399-2:1997), (EN 1991-1-4), (IS-875 Part 3) to estimate the design pressure coefficients and force coefficients for such buildings subjected to wind induced loads. CFD has expanded as a tool to replace wind tunnel testing as it is quicker, less expansive and gives more information and control to designers. CFD for determining wind responses is becoming immensely popular as the analysis saves time in the design process and is hence cheaper and less time-consuming compared to conventional testing. Since the 1960s, several researchers have derived a wealth of useful information by assuming that wind pressure behaves in a static manner. Urbanization in the twentieth and twenty-first centuries stimulated the development of tall buildings, which became taller and taller as demand increased. (Amin & Ahuja, 2013) summarized the outcomes of wind tunnel tests on 1:300 scaled down models of rectangular buildings having the same height and plan area but different side ratios. They concluded that the side ratio of models has a significant impact on the amplitude and distribution of wind pressure on leeward and sidewalls, but only a little impact on windward walls at a 0-degree wind incidence angle. (Revuz et al., 2012) investigated the size of the domain size for conducting CFD in the field of wind engineering. This paper discussed what size of domain to use while using CFD tools for simulation of tall buildings. (Kastner & Dogan, 2019) estimated the time-saving

2nd International Conference on Futuristic and Sustainable Aspects in Engineering and Technology
AIP Conf. Proc. 2721, 060002-1–060002-16; https://doi.org/10.1063/5.0154002
Published by AIP Publishing. 978-0-7354-4576-5/$30.00

methodology that is possible along with specific mesh properties to take the lead in the proposed method. In this research study, they proposed a cylindrical meshing for wind simulations. They tried to compare the box-shaped simulation domain with the cylindrical computational domain, and showed that it is recommended to use the box-shaped simulation approach if no annual wind analysis is intended. Franke et al. (2004) reviewed the outcomes from published simulations and attempted to make a recommendation on how to use CFD for wind engineering tasks, focusing on the statistically steady simulation of wind. The results for mean velocities and turbulence in the constructed environment were summarized from statistically steady RANS simulations accessible in the literature, and advice on how to use CWE properly for that purpose was given. Huang and Chen (2007) attempted to evaluate the wind load response and equivalent static wind loads (ESWLs) of high-rise buildings based upon measured synchronic surface pressures in a wind tunnel procedure. They proposed a practical procedure for modelling ESWLs on tall buildings, linking the effect of higher mode contributions. Li et al. (2007) used wind tunnel tests and full-scale field measurements to explore the effect of wind on the Jin Mao Building under strong winds. Mallick et al. (2018) studied the simulation of the wind pressure coefficient on C-shaped building models by means of numerical analysis using ANSYS Fluent and concluded that the pressure on the building was remarkably influenced by the structure geometry, orientation, aspect ratios, and wind angle of attack. Fouad et al. (2018) compared pressure coefficients from ASCE code, Euro code and CFD simulation results for low-rise buildings and concluded that the CFD simulation findings are substantially identical to the experimentally determined values for the walls and leeward side of the roof; however, there is a difference in the results for the windward side of the roof. Kwon & Kareem (2013) investigate the similarities and differences in the wind response of tall buildings based upon major international wind standards such as ASCE, AS/NZ, AIJ, CNS, NBCC, EU, ISO, and IWC and a database-enabled design procedure, Nat-Haz Aerodynamic Loads Database (NALD). (Nagar et al., 2020) investigated H-shaped and square-shaped buildings under isolated as well as interference conditions due to surrounding buildings. They examined the mean pressure coefficient using wind tunnel experiments for various angles of attack of the wind. (Pal et al., 2021) discussed twin square and fish-plan form building models with equal volume for wind-induced mutual interference effects and investigated the maximum efficiency in terms of wind pressure and base shear for both the shapes and concluded for the most efficient condition. Bhattacharyya & Dalui (2018) presented a detailed study on an E-plan-shaped building that is asymmetrical about both the plan axes under wind excitation. The angle of incidence of the wind varied from 0° to 330° at an interval of 30°. Assainar & Dalui (2020) discussed the effectiveness of aerodynamic alterations performed on a pentagonal-plan shaped model using computational fluid dynamics (CFD) simulation. The study focused on evaluating the effectiveness of different corner modifications applied to a pentagon-shaped tall building. Sanyal & Dalui (2020a) investigated the effect of major shape modifications on a Y-plan-shaped building was studied statistically by computational fluid dynamics. Sanyal & Dalui (2020b) studied the wind pressure distribution on a Y-shaped tall building owing to changes in the building side ratio where the length-to-width ratio was altered while the total plan area remained the same for all models. Tominaga et al. (2008) evaluated computational fluid dynamics (CFD) findings for flow around a high-rise building model with a 1:1:2 shape placed within the surface boundary layer using several revised k-models and large eddy simulation (LES).

10 August 2023 06:50:22

NUMERICAL ANALYSIS

In this study, rectangular-plan shaped buildings with different side ratios are analyzed by using the ANSYS CFX package for CFD simulation.

Governing Equation

The model's mean flow characteristics are simulated using a κ-ε turbulence model. This model uses two equations by means of two transport equations. For the turbulence model used, instantaneous velocity can be written as the summation of time-averaged mean velocity and a time-varying fluctuating component given in equation (1).

$$u_i = U_i + u_i' \tag{1}$$

where, u_i = instantaneous velocity

U_i = time averaged mean velocity

u_i' = fluctuating component of velocity

As per Reynold's Average Navier Stokes Equation (RANS) equation (2)

$$\frac{\partial}{\partial x_i}(\rho U_i) = 0 \tag{2}$$

where, ρ = density of fluid

and conservation of momentum equation (3)

$$\frac{\partial}{\partial x_j}(\rho U_i U_j) = -\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_f}\left[\mu\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_f}{\partial x_i}\right)\right] + \frac{\partial}{\partial x_j}\left(-\rho\overline{u_i' u_j'}\right) \tag{3}$$

Building Models

Rectangular building model with different side ratio are considered in this study whose plan shape and elevation are as shown in fig.1 and fig.2 respectively. All the models analyzed are of same plan area 900 m2 and height 90 m modelled on a scale of 1:100 tabulated as in Table 1.

Table 1 Model Dimensions

Modal shape	Prototype Dimension(m)	ANSYS Model Dimension(mm)	Height of Model (mm)	Scale of Model
Square	30*30	300*300	900	1:100
Rectangle-1	36*25	360*250		
Rectangle-2	45*20	450*200		
Rectangle-3	60*15	600*150		

The models are rotated from 0° to 90° at 30° interval. Therefore, attacking wind angles are 0°, 30°, 60°, 90°. This study also validates with wind pressure of an isolated square plan shape, tall building using CFD simulation and compared it with wind effects of conventional plan shape building given in Indian code IS: 875 (Part- 3), 2015.

Domain and Meshing

In the case of high-rise buildings, domain size is mainly governed by the height of the building, such that a large number of cell counts could be formed and many of them used up in the region far away from the wake region. The domain size selected in modelling is defined as per Zidan et al. (2004). The inlet and outlet distances of the domain from the building position are taken as 4H and 15H, respectively. The side aspect and top clearance are also taken as 5H and 8H, where H is the height of the building. The domain configurations are depicted in Fig 3.

10 August 2023 06:50:22

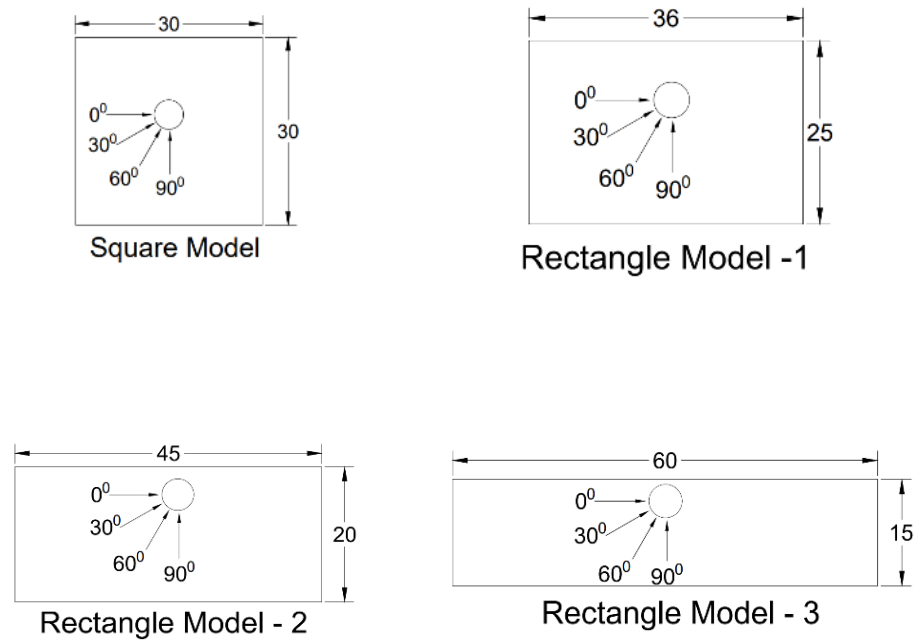


Fig 1. Plan Dimension of Models

.

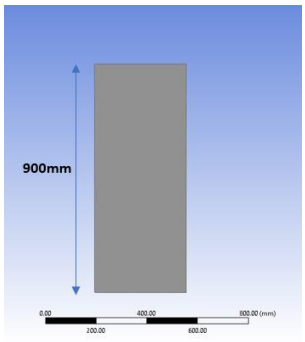


Fig 2. Elevation of Model

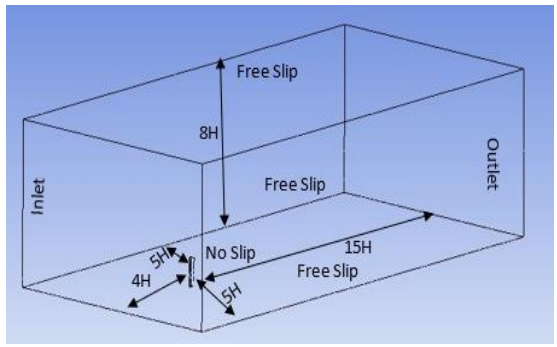


Fig 3. Modelling of Tunnel

The meshing of buildings is in fig. 4, and domain meshing is in fig. 5. Domain meshing is done by tetrahedral mesh elements. Meshing near the buildings is made comparatively finer to enhance the accuracy of the results. The inlet velocity is taken as 10 m/s.

10 August 2023 06:50:22

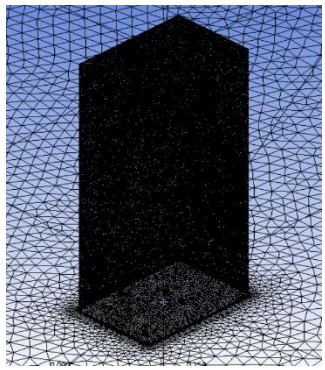


Fig 4. Building Meshing

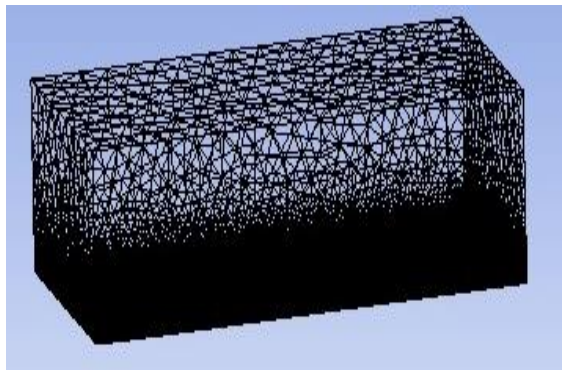


Fig 5. Domain Meshing

Velocity Profile and Turbulent Intensity Profile

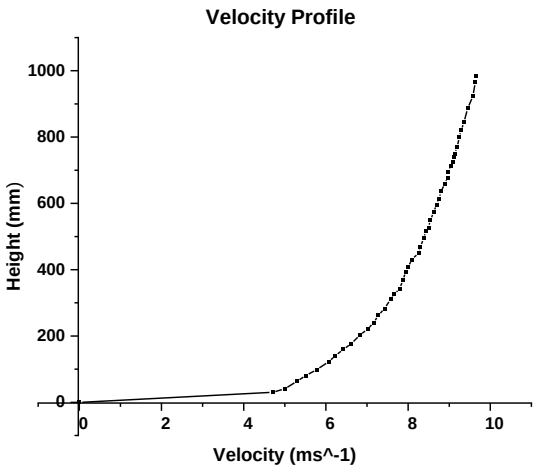


Fig 6. Variation of velocity with height

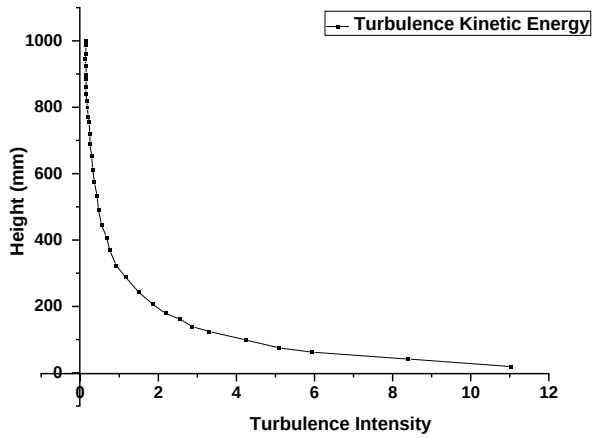


Fig 7. Variation of turbulence intensity with height

The wind profile boundary layer is governed by the power law equation. As per Power's Law, the wind speed profile within the atmospheric boundary layer is given by:

$$\frac{V}{Vo} = \left[\left(\frac{Z}{Zo}\right) \wedge \left(\frac{1}{alpha}\right)\right] \quad (4)$$

Where, V = velocity of wind at height Z

Vo = gradient velocity of wind at reference height Zo

Z = height above ground

Zo = Nominal height of Boundary layer (also called gradient height)

alpha = power law coefficient.

10 August 2023 06:50:22

Validation

The validity of the ANSYS CFX software is validated before beginning the numerical study of the building. For this purpose, a square-plan shaped building with dimensions of 100 mm 100 mm and a height of 500 mm (i.e., aspect ratio of 1:5) is investigated using the k-epsilon model with ANSYS CFX in the domain under uniform wind flow. The ANSYS CFX programme determines the face average values of coefficient of pressure, which are then compared to wind action codes from various regions. The velocity profile in fig. 6 and the turbulent kinetic energy curves so obtained are as shown in fig. 7.

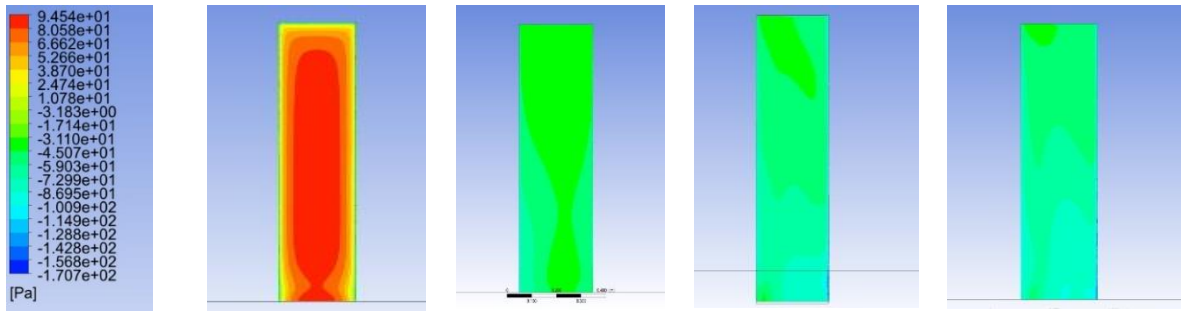
Results and Discussion

Result which are investigated after investigating the wind generated effects are investigated for the building and results are explain in the following section.

Pressure Distribution

As below, pressure distribution for various faces of a building model is represented using contour plots as below:

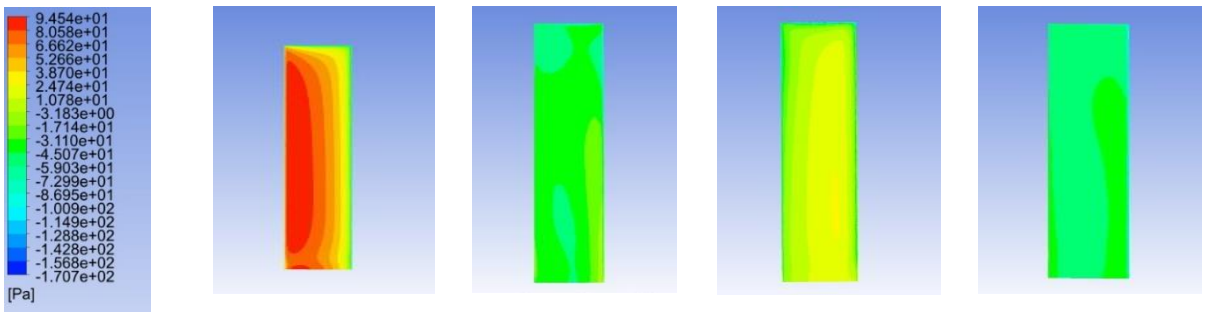
Model- 1 Square shaped Building 0^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 8. Contour Plots for different faces at 0-degree of incidence

Model- 1 Square shaped Building 30^0 Contour

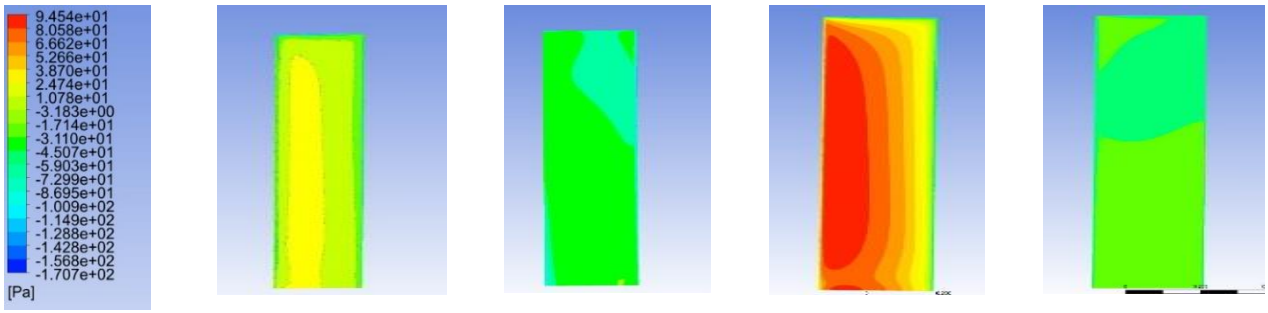


Legend **Face A** **Face B** **Face C** **Face D**

Fig 9. Contour Plots for different faces at 30-degree of incidence

10 August 2023 06:50:22

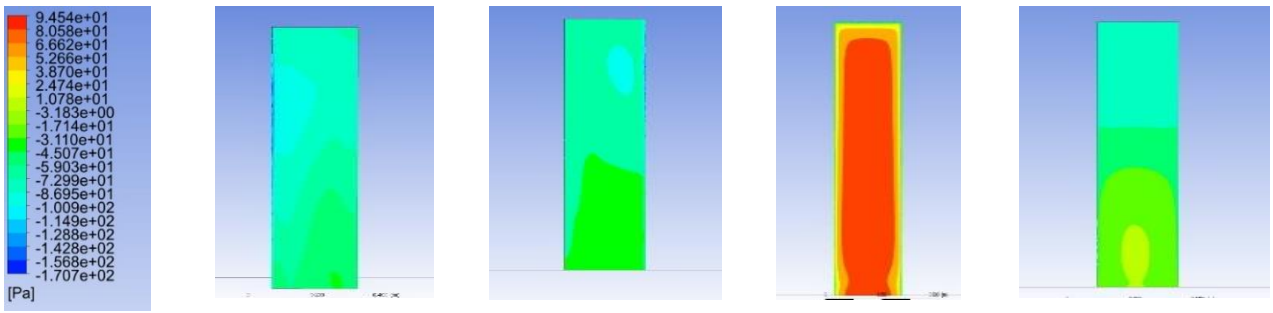
Model- 1 Square shaped Building 60^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 10. Contour Plots for different faces at 60-degree of incidence

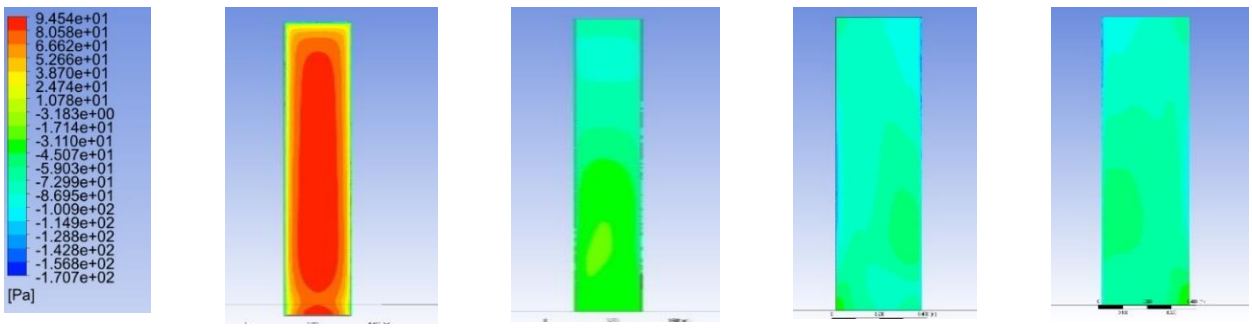
Model- 1 Square shaped Building 90^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 11. Contour Plots for different faces at 90-degree of incidence

Model 2 Rectangular shape building with side ratio of 1.44 at 0^0 Contour

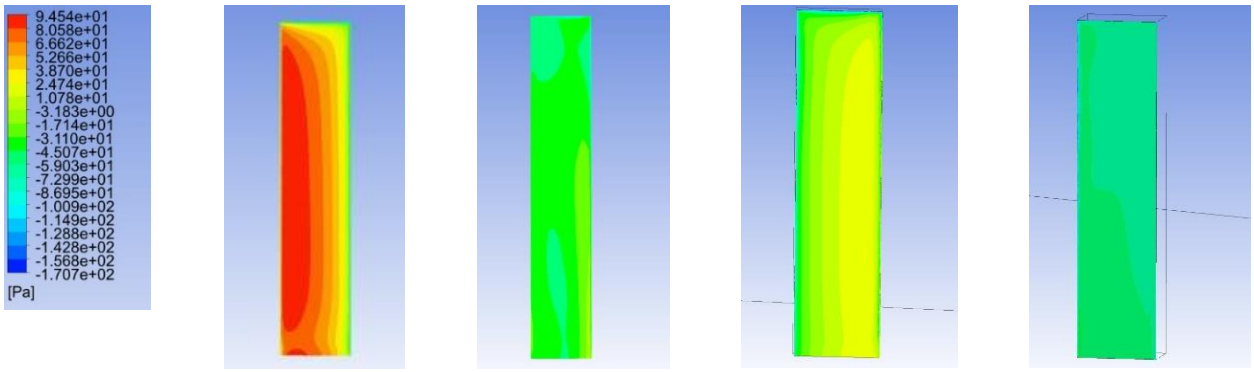


Legend **Face A** **Face B** **Face C** **Face D**

Fig 12. Contour Plots for different faces at 90-degree of incidence

10 August 2023 06:50:22

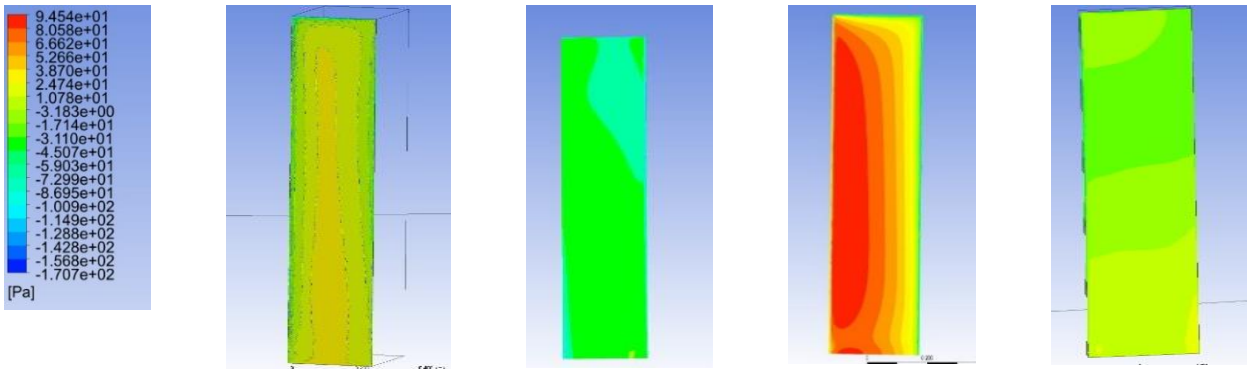
Model 2 Rectangular shape building with side ratio of 1.44 at 30^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 13. Contour Plots for different faces at 90-degree of incidence

Model 2 Rectangular shape building with side ratio of 1.44 at 60^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 14. Contour Plots for different faces at 60-degree of incidence

• The pressure contours for different faces at various angles of incidence depict the pressure distribution on faces and are shown in figures.

• For Model 1, contour plots for various faces showing pressure distribution with different angles of wind incidence have been shown as in fig. 8 to fig. 11.

• Initially, at 0 degree of incidence, Face A shows a positive pressure being the windward face whereas face B, C, and D depicts a negative pressure distribution being the leeward and side wall face of the building.

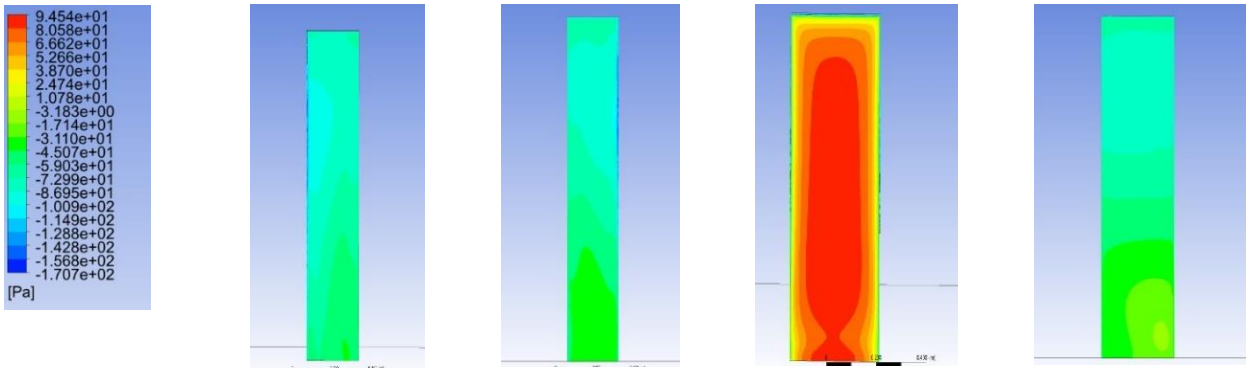
• As the angle of wind incidence changes, the pressure distribution changed showing negative pressure at face A and face C showing positive pressure at 90-degree wind incidence.

• For Model 2, contour plots for different faces showing pressure distribution with different angles of wind incidence have been shown as in figs. 12 to 15.

10 August 2023 06:50:22

• For Model 3 and Model 4, contour plots for various faces with different angles of wind incidence have been shown as in figs. 16-19 and 20-23.

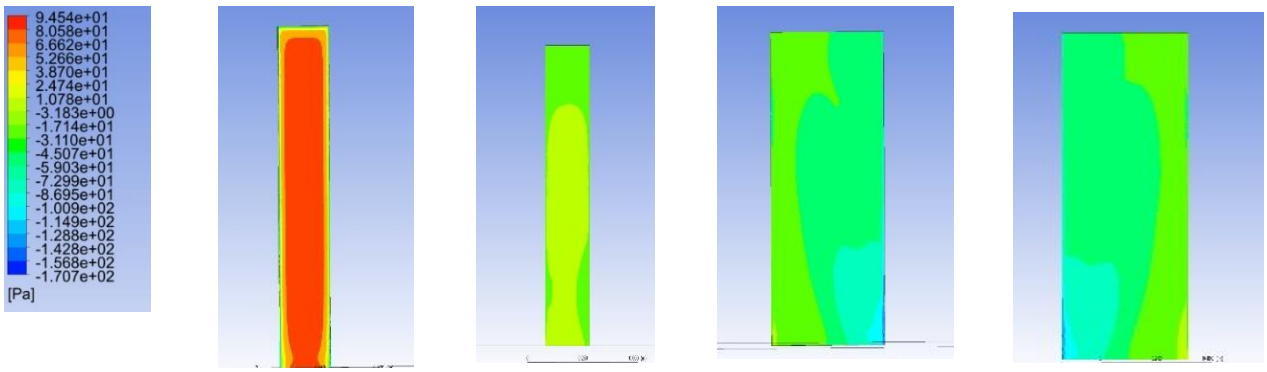
Model 2 Rectangular shape building with side ratio of 1.44 at 90^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 15. Contour Plots for different faces at 90-degree of incidence

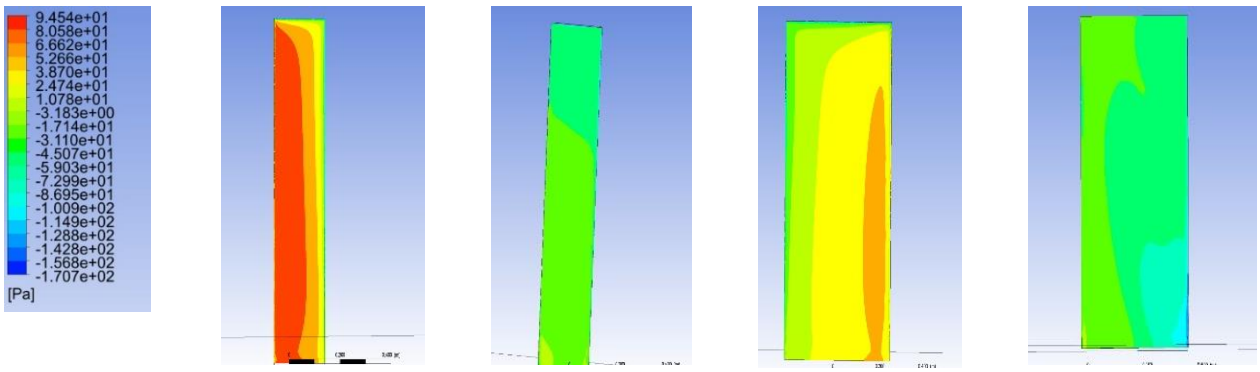
Model 3 Rectangular shape building with side ratio of 2.25 at 0^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig16. Contour Plots for different faces at 0-degree of incidence

Model 3 Rectangular shape building with side ratio of 2.25 at 30^0 Contour

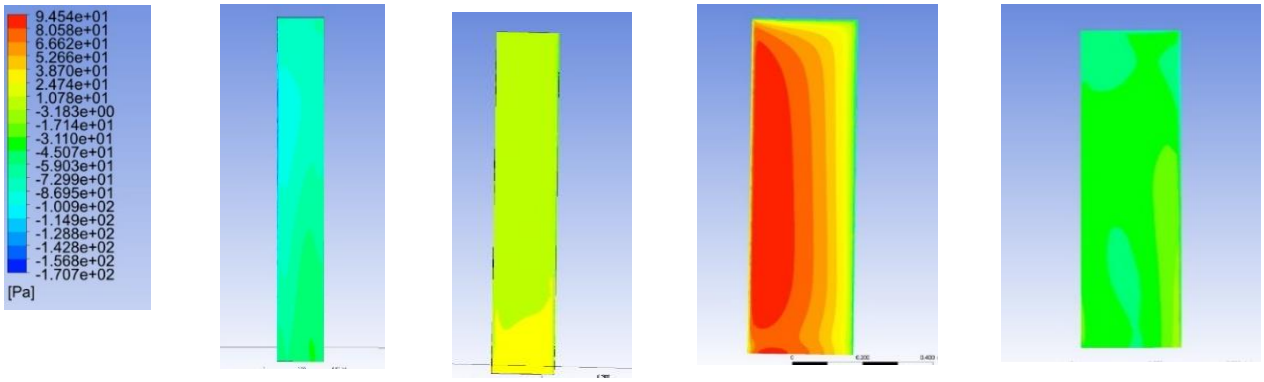


Legend **Face A** **Face B** **Face C** **Face D**

Fig 17. Contour Plots for different faces at 30-degree of incidence

10 August 2023 06:50:22

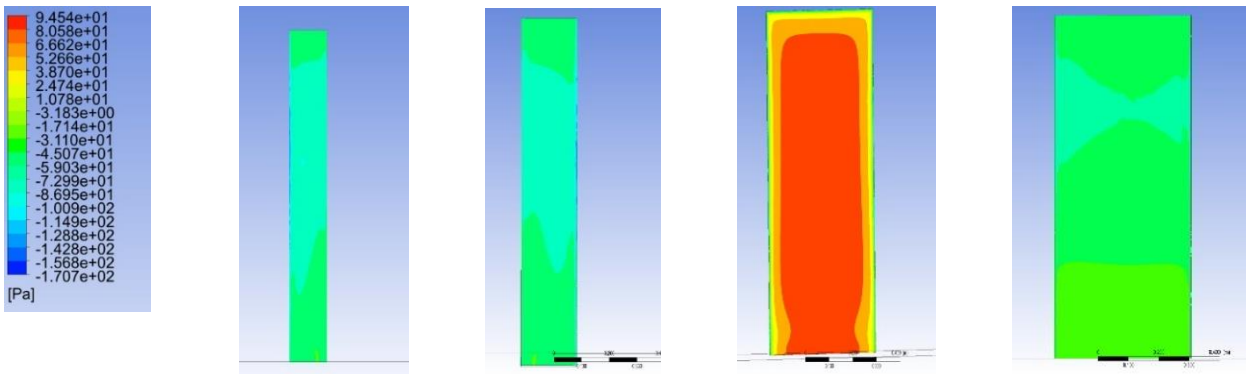
Model 3 Rectangular shape building with side ratio of 2.25 at 60^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 18. Contour Plots for different faces at 60-degree of incidence

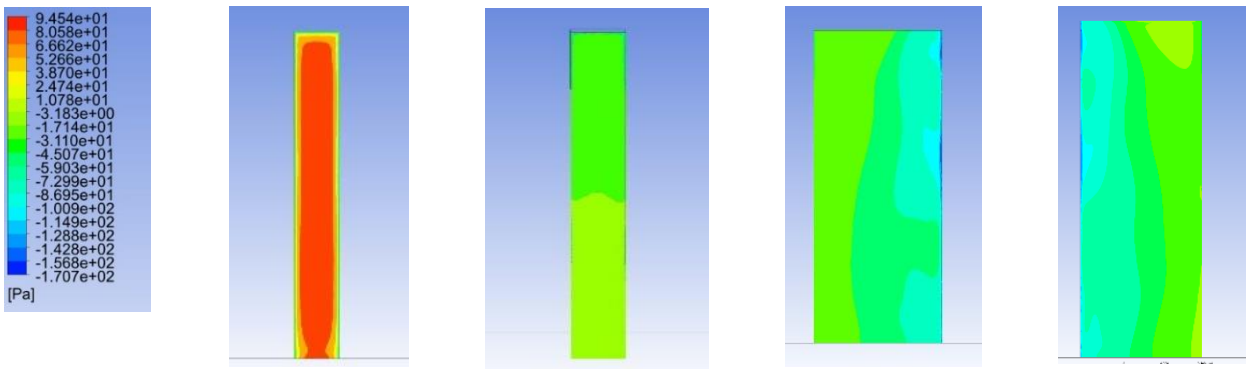
Model 3 Rectangular shape building with side ratio of 2.25 at 0^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 19. Contour Plots for different faces at 60-degree of incidence

Model 4 Rectangular shape building with side ratio of 2.25 at 0^0 Contour

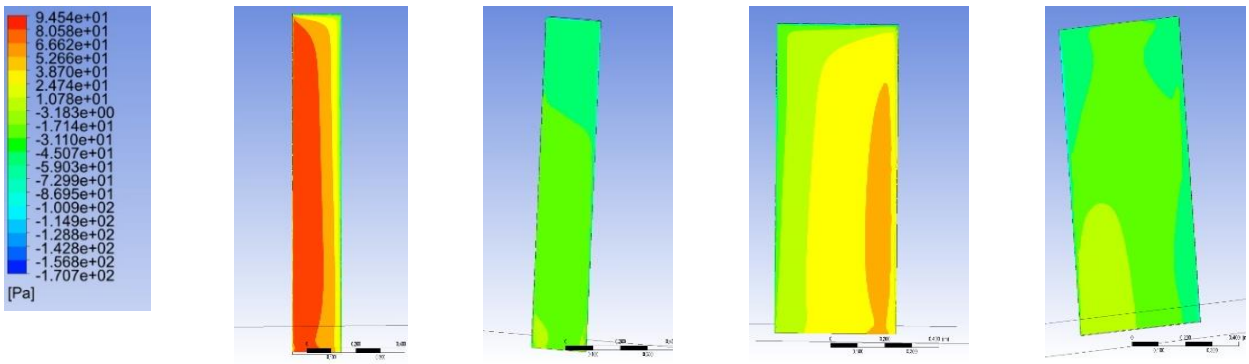


Legend **Face A** **Face B** **Face C** **Face D**

Fig 20. Contour Plots for different faces at 0-degree of incidence

10 August 2023 06:50:22

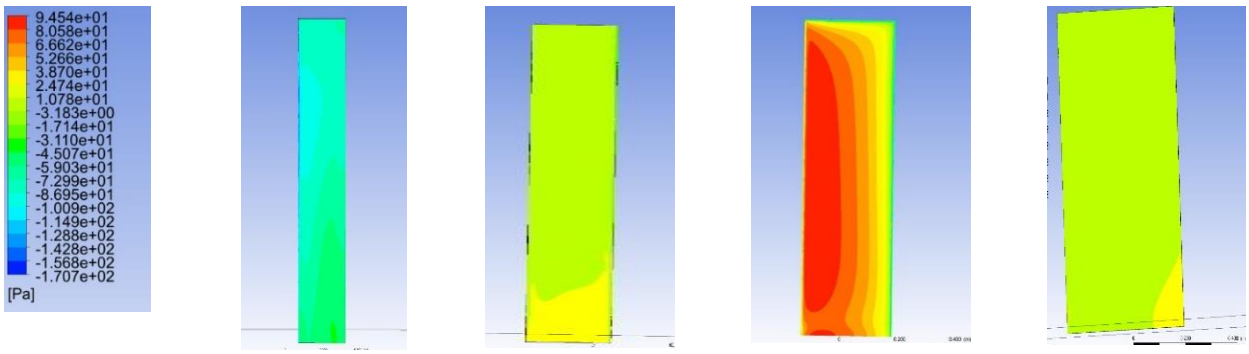
Model 4 Rectangular shape building with side ratio of 2.25 at 30^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 21. Contour Plots for different faces at 30-degree of incidence

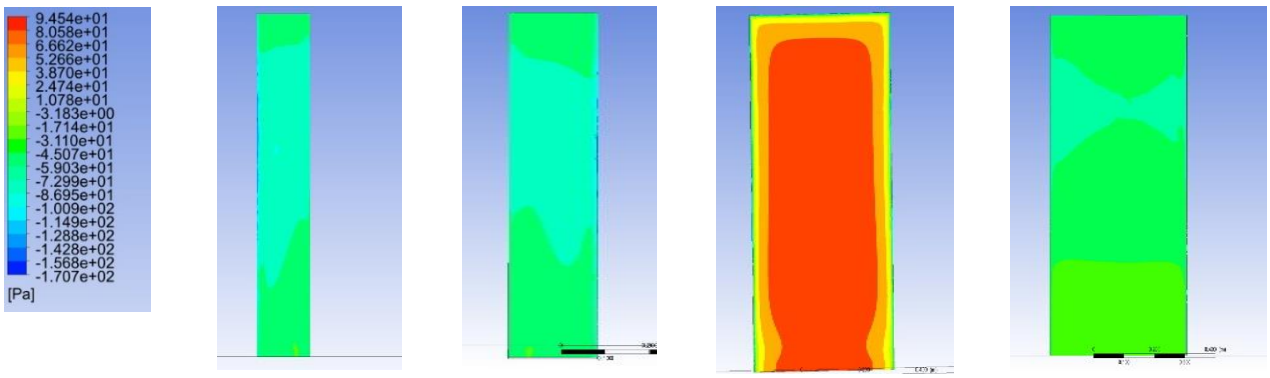
Model 4 Rectangular shape building with side ratio of 2.25 at 60^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 22. Contour Plots for different faces at 60-degree of incidence

Model 4 Rectangular shape building with side ratio of 2.25 at 90^0 Contour



Legend **Face A** **Face B** **Face C** **Face D**

Fig 23. Contour Plots for different faces at 90-degree of incidence

10 August 2023 06:50:22

Mean Pressure Coefficient:

The mean pressure coefficient '$C_{p\ mean}$' is calculated from the equation (1)

$$C_{p\ mean} = \frac{(p - p_o)}{\frac{1}{2}\rho U_H^2} \quad (5)$$

where p is the pressure at point on surface,

p_o is the reference height static pressure,

ρ is the air density

U_H is mean wind velocity at the building reference height.

The external pressure coefficient is calculated by using the expression,

$$C_{pe} = \frac{p}{0.6 \times V_z^2} \quad (6)$$

The mean pressure coefficient for different faces is evaluated and compared with the standards as shown in fig 24.

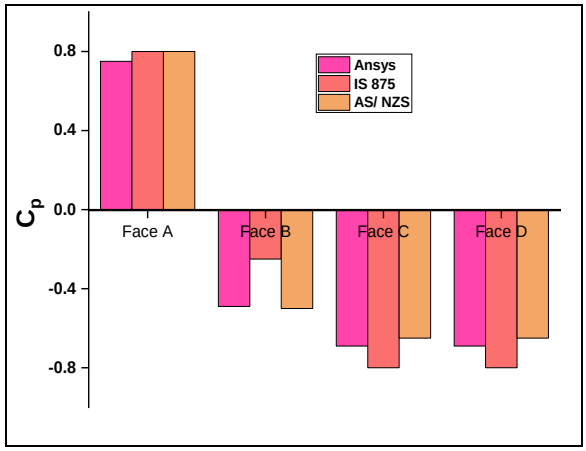


Fig 24. Comparison of mean pressure coefficients on a different face of all model for 0-degree between numerical result and International Standards

Vertical Centreline Pressure Coefficients

The curve in these figures depicts the variation in the pressure coefficient Cp value for various faces as a function of height. Therefore, the face average value of Cp for several models at various angles of incidence is evaluated. The vertical centreline Cp variation with height for Model 1, Model 2, Model 3, and Model 4 is plotted as shown in figs. 25, 26, 27, and 28.

10 August 2023 06:50:22

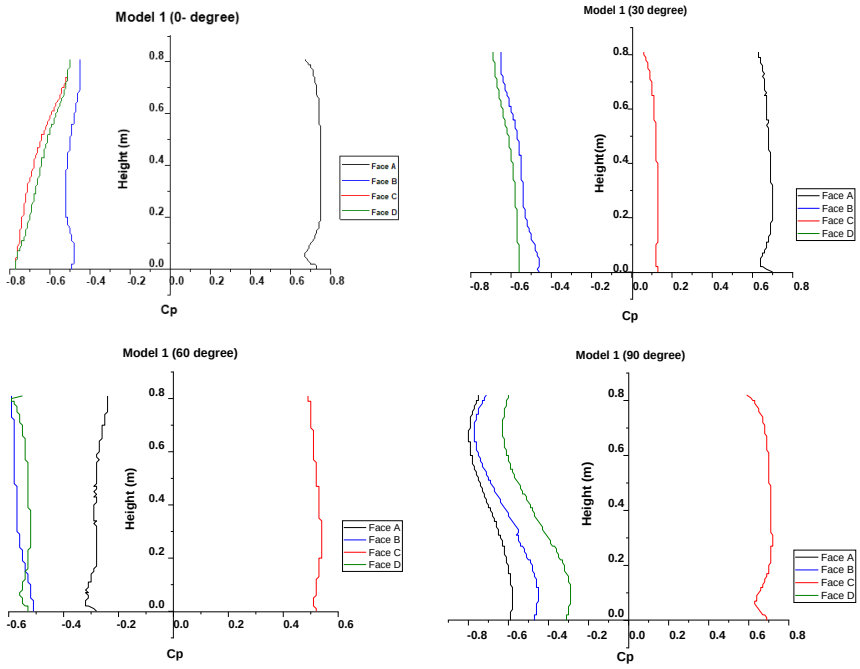


Fig 25. Pressure Variation along Centerline for all faces for different degree AOA of model 1

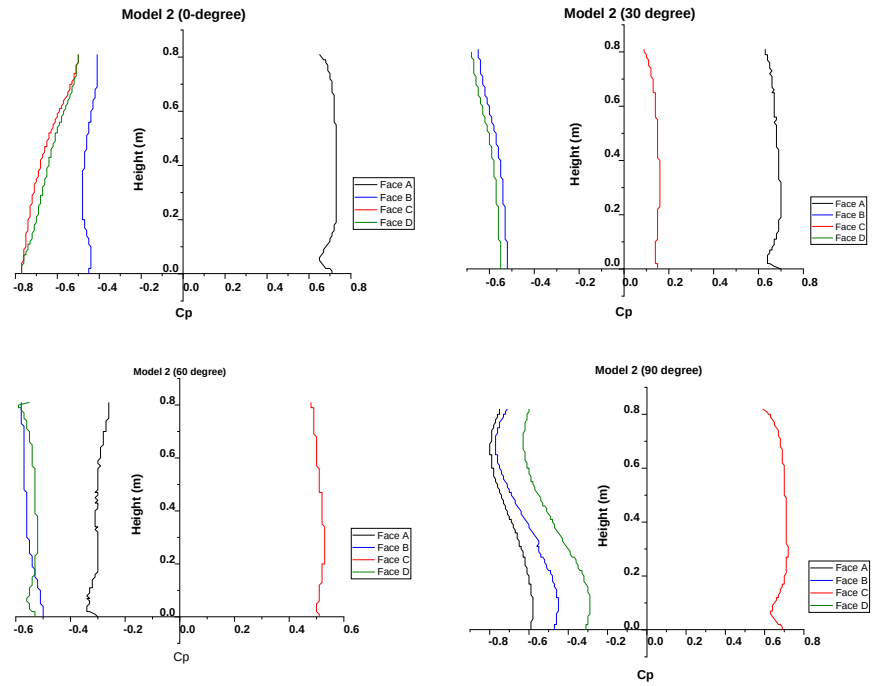


Fig 26. Pressure Variation along Centerline for all faces for different degree AOA of model 2

10 August 2023 06:50:22

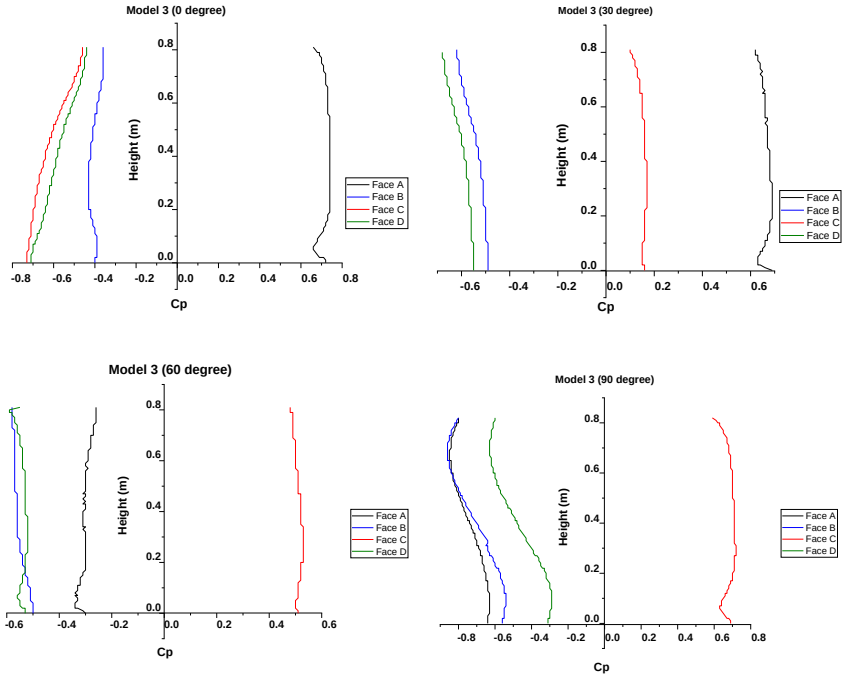


Fig 27. Pressure Variation along Centerline for all faces for different degree AOA of model 3

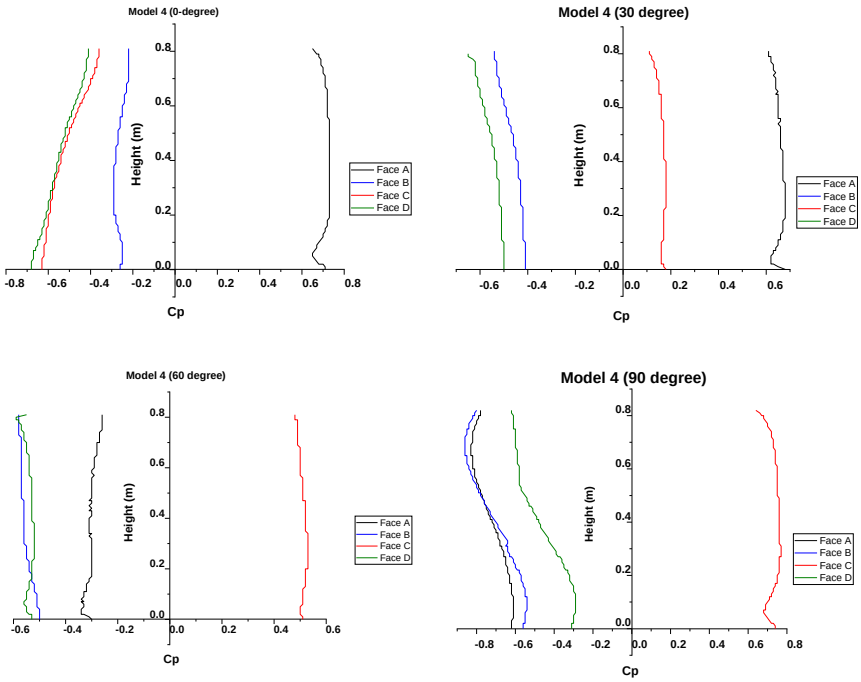


Fig 28. Pressure Variation along Centerline for all faces for different degree AOA of model 4

10 August 2023 06:50:22

CONCLUSION

The pressure contour and mean pressure coefficients for the rectangular-shape building models for different side ratios at 0^0 ,30^0 ,60^0 and 90^0 wind incidence angles are compared in this paper. The k-ε model is used to simulate the results. The major outcome of this research is as below:

- The effect of side ratios and wind directions on wind pressure distribution and the magnitude of pressure coefficients on rectangular building models is identified by a numerical study measurement of wind pressures on building models.
- The variation of pressure coefficients on the centreline is discussed and plotted graphically.
- The model side ratio has a significant impact on the amplitude and distribution of wind pressure on leeward and sidewalls, but only a little impact on windward walls when the wind incidence angle is zero.
- At 0 degrees of incidence, face A portrays a positive pressure distribution as the building's windward face, whereas faces B, C, and D depict a negative pressure.
- As the angle of incidence changes, the pressure distribution changes, and so does the pressure coefficient. For 30° and 60° angles of incidence, positive pressure distribution at face A becomes slightly less than in the case of 0°, along with suction pressure increase at face D, and a comparable change can be viewed from the pressure coefficient data so obtained as a result of the distribution as the building's leeward and sidewall faces.
- At 90° angle of incidence, face C becomes windward and a similar contour plot pattern as that of face A when the angle of attack was 0° can be seen.
- The present study shows that changes in side ratio have considerably affected the general magnitude of peak pressures and peak suctions in models with constant cross section.

10 August 2023 06:50:22

ACKNOWLEDGEMENT

The authors are thankful to **Delhi Technological University** for providing the research facilities and funding to support this study. authors would like to express their appreciation to Manvendra Verma, Indrajeet Singh for their precious help with the numerical simulation and to DTU civil engineering faculty, who encouraged and supported the authors during the entire study.

REFERENCES

[1] Amin, J. A., & Ahuja, A. K. (2013). Effects of Side Ratio on Wind-Induced Pressure Distribution on Rectangular Buildings. Journal of Structures, 2013, 1–12. https://doi.org/10.1155/2013/176739

[2] AS/NZS 1170.2:2011 :: Standards New Zealand. (n.d.). Retrieved July 17, 2021, from https://www.standards.govt.nz/shop/asnzs-1170-22011/

[3] ASCE 7-10: Minimum Design Loads for Buildings and Other Structures. (n.d.).

[4] Assainar, N., & Dalui, S. K. (2020). Aerodynamic analysis of pentagon-shaped tall buildings. Asian Journal of Civil Engineering 2020 22:1, 22(1), 33–48. https://doi.org/10.1007/S42107-020-00296-2

[5] Bhattacharyya, B., & Dalui, S. K. (2018). Investigation of mean wind pressures on 'E' plan shaped tall building. Wind and Structures, An International Journal, 26(2), 99–114. https://doi.org/10.12989/WAS.2018.26.2.099

[6] BS 6399-2:1997 - Loading for buildings. Code of practice for wind loads (British Standard). (n.d.). Retrieved July 17, 2021, from https://webstore.ansi.org/Standards/BSI/BS63991997

[7] EN 1991-1-4: Eurocode 1: Actions on structures - Part 1-4: General actions - Wind actions. (2010).

[8] Fouad, N. S., Mahmoud, G. H., & Nasr, N. E. (2018). Comparative study of international codes wind loads and CFD results for low rise buildings. Alexandria Engineering Journal, 57(4), 3623–3639. https://doi.org/10.1016/J.AEJ.2017.11.023

[9] Franke, J., Hirsch, C., Jensen, A. G., Krus, H. W., Schatzmann, P. S., Miles, S. D., Wisse, J. A., & Wright, N. G. (2004). Recommendations on the use of CFD in wind engineering. COST Action C14: Impact of Wind and Storm on City Life and Urban Environment. https://www.researchgate.net/publication/251814717_Recommendations_on_the_use_of_CFD_in_wind_engineering

[10] Huang, G., & Chen, X. (2007). Wind load effects and equivalent static wind loads of tall buildings based on synchronous pressure measurements. Engineering Structures, 29(10), 2641–2653. https://doi.org/10.1016/J.ENGSTRUCT.2007.01.011

[11] IS 875.3: Code of Practice for Design Loads (Other than Earthquake) for Buildings and Structures - Part 3 : Wind Loads (IS 875 : Part 3) + Amendment 2016 : Bureau of Indian Standards : Free Download, Borrow, and Streaming : Internet Archive. (n.d.). Retrieved July 17, 2021, from https://archive.org/details/gov.in.is.875.3.2015/page/n3/mode/2up

[12] Kastner, P., & Dogan, T. (2019). A cylindrical meshing methodology for annual urban computational fluid dynamics simulations. Https://Doi.Org/10.1080/19401493.2019.1692906, 13(1), 59–68. https://doi.org/10.1080/19401493.2019.1692906

[13] Li, Q. S., Xiao, Y. Q., Fu, J. Y., & Li, Z. N. (2007). Full-scale measurements of wind effects on the Jin Mao building. Journal of Wind Engineering and Industrial Aerodynamics, 95(6), 445–466. https://doi.org/10.1016/J.JWEIA.2006.09.002

[14] Mallick, M., Mohanta, A., Kumar, A., & Raj, V. (2018). Modelling of Wind Pressure Coefficients on C-Shaped Building Models. Modelling and Simulation in Engineering, 2018. https://doi.org/10.1155/2018/6524945

[15] Nagar, S. K., Raj, R., & Dev, N. (2020). Experimental study of wind-induced pressures on tall buildings of different shapes. Wind and Structures, An International Journal, 31(5), 441–453. https://doi.org/10.12989/was.2020.31.5.431

[16] Pal, S., Raj, R., & Anbukumar, S. (2021). Comparative study of wind induced mutual interference effects on square and fish-plan shape tall buildings. Sādhanā 2021 46:2, 46(2), 1–27. https://doi.org/10.1007/S12046-021-01592-6

[17] Revuz, J., Hargreaves, D. M., & Owen, J. S. (2012). On the domain size for the steady-state CFD modelling of a tall building. Wind and Structures, An International Journal, 15(4), 313–329. https://doi.org/10.12989/WAS.2012.15.4.313

[18] Sanyal, P., & Dalui, S. K. (2020a). Effect of corner modifications on Y' plan shaped tall building under wind load. Wind and Structures, An International Journal, 30(3), 245–260. https://doi.org/10.12989/was.2020.30.3.245

[19] Sanyal, P., & Dalui, S. K. (2020b). Effects of side ratio for 'Y' plan shaped tall building under wind load. Building Simulation 2020 14:4, 14(4), 1221–1236. https://doi.org/10.1007/S12273-020-0731-1

[20] Tominaga, Y., Mochida, A., Murakami, S., & Sawaki, S. (2008). Comparison of various revised k–ε models and LES applied to flow around a high-rise building model with 1:1:2 shape placed within the surface boundary layer. Journal of Wind Engineering and Industrial Aerodynamics, 96(4), 389–411. https://doi.org/10.1016/J.JWEIA.2008.01.004

10 August 2023 06:50:22

Proceedings of the 5th International Conference on Inventive Research in Computing Applications (ICIRCA 2023)
IEEE Xplore Part Number: CFP23N67-ART; ISBN: 979-8-3503-2142-5

An Overview on Integration of Artificial Intelligence Tools to Predict the Nature of Chronic Obstructive Pulmonary Disease

Sourabh Sharma
Department of Biotechnology
Delhi Technological University
New Delhi, India
sourabhsharma2598@gmail.com

Yasha Hasija*
Department. of Biotechnology
Delhi Technological University
New Delhi, India
yashahasija06@gmail.com
*Corresponding Author

2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) | 979-8-3503-2142-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIRCA57980.2023.10220762

***Abstract*— Artificial intelligence (AI) and its nature have evolved over the period of time. With its rapid growth and application, it has shown many remarkable changes, especially in the identification and diagnosis of various diseases including lung disorders. One such disease is Chronic obstructive pulmonary disease (COPD). COPD is a chronic lung condition that makes breathing difficult. It is characterized by a prolonged airflow blockage that interferes with normal breathing. The high amount of data including various test reports and images sample that can be obtained on the clinical level and seeing the nature of COPD makes this disease an ideal subject that can be uncovered using the help of many AI and ML tools. This review focuses on the application of ML models and AI tools that have played a vital role in suggesting the nature of COPD including the areas where the tools have limitations and the further opportunities where the tools can be improved.**

Keywords—Artificial Intelligence; Machine learning; Deep Learning; Chronic obstructive pulmonary disease; Emphysema; Respiratory disease

I. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a steady developing, incurable lung condition marked by a persistent airflow restriction that leads to more widespread illnesses such as respiratory failure and pulmonary heart disease. Prolonged obstruction of the lungs' airways is what leads to COPD. This obstruction has a significant negative impact on the heart's ability to pump blood and on gas exchange, both of which have systemic repercussions. Emphysema, where the pulmonary air sacs get injured, or chronic bronchitis, which causes ongoing airway inflammation, are the two main causes of this illness. In addition to heart failure, bronchiectasis, cardiovascular illnesses, sleep apnea, asthma, frailty, and malnutrition are some serious symptoms in COPD patients, also these patients frequently suffer from other pulmonary and systemic comorbidities [1].

Despite significant efforts, there are still many obstacles in the diagnosis, management, and treatment of this disorder these include the absence of consistent classification criteria for different phenotypes, an unclear understanding of its pathogenesis, and the substantial costs associated with exacerbations. Additionally, COPD carries a high mortality risk. Overcoming these challenges is crucial to improve outcomes for individuals with COPD, necessitating the development of standardized criteria, enhanced understanding of disease mechanisms, cost-effective strategies to manage exacerbations, and interventions to reduce mortality rates.

The way we interpret and analyze complex systems using clinical, imaging, and molecular data has been transformed by artificial intelligence (AI). This led to the growth of various AI applications in different fields, including healthcare. Despite the fact that applied AI is still an emerging concept in numerous disciplines, this has demonstrated intriguing and significant outcomes when used to diagnose various ailments using primary sources of data such as diagnostic images [2].

AI includes machine learning (ML), which uses statistical techniques for self-learning without explicit programming. Machine learning is frequently compared to a "black box," where input goes into one side and a solution is displayed on the other. It is particularly true for neural networks, where data might go through intricate algorithmic transformations involving many layers, and the model may be highly complicated and uncertain to work with.

Recently, a number of AI/ML techniques have been used for COPD, but only a small number have made a substantial impact on clinical practice. In order to use AI/ML tools safely and effectively, it is necessary to summarize current knowledge and suggest future paths in the field of computational biology.

II. ADVANCEMENT OF MEDICINE USING AI

Machine learning can be helpful since it enables important associations to be drawn from data between different pieces of information instead of needing a specific human definition. The ability of a technique based on ML to consider a larger variety of data may result in a more accurate diagnosis. Furthermore, defining the correlations within the dataset will enable a better understanding of the condition. ML is additionally will play a key part in enabling medical professionals to provide better management of diseases in patients, from a scientific or medical perspective [3]. There are many advantages of applying AI, one of them being it makes it easy to observe data such as medical images and tests obtained from underserved areas or remote areas. This helps in the early identification of diseases and to send the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:44 UTC from IEEE Xplore. Restrictions apply.

patient to expert care on time, if necessary, which could redirect the course of the disease's outcome.

Using AI, in 2018 a medical device by US Food and Drug Administration (FDA) was accepted to identify the levels of diabetic retinopathy in patients having diabetes. The device gives a screening conclusion observation that does not require a specialist to evaluate the outcome, making it useful for healthcare professionals that might not typically be taking part in eye care and allowing patients to receive an initial diagnosis and severity determination without having to wait to be referred to specialists.

TABLE I. LIMITATIONS AND ISSUES IN EXISTING MODELS

CRITERIA	DESCRIPTION
Inadequate capability of current procedures	The article states that the current procedures for categorizing COPD patients into phenotypes are insufficient and require additional information
Limited dataset representation	Some studies mentioned in the article had limitations in terms of the percentage of female participants, which could affect the generalizability of the results.
Dependency on specific tests	Some AI/ML models relied on lung function tests to assess hindrance in airflow, but not all patients were able to complete these tests, leading to missing data.
Dataset size and diversity:	The datasets used in some studies were relatively small or limited to specific ranges of COPD severity, which could affect the accuracy and generalizability of the models.
Training time and computational resources:	Certain approaches mentioned in the article required significant training time and computational resources, making them less practical for widespread use.
Scope and additional features:	The applicability of some models was limited to specific contexts, such as hospitalized patients or acute exacerbations of COPD. The article suggests that including additional features in the dataset could enhance the models' scope and performance.

A supporting tool for a clinical conclusion that examines the brain's computed tomography (CT) images for signs of a stroke and alerts a neurovascular expert if a potential large blockage in the vessel is found was approved by the FDA in 2018. Instead of requiring a neurovascular specialist to be notified once a radiologist examines the CT images, the algorithm automatically alerts the specialist while the first-line healthcare professional is performing a regular assessment of the images. Patients thus, receive the best care sooner for an illness where a positive result is highly time-sensitive.

Even so, the healthcare systems have advanced their delivery mechanisms to make sure the secure distribution of drugs to individuals, [4] The various application of ML will necessitate a similar polished network of governing oversight, local norms, and valid framework to assure the safe usage, monitoring and development of AI systems [4]. Additionally, scalable computer infrastructure will be needed to handle the massive volumes of data generated by the use of these models. In order to advance the field and reduce the unfavorable effects of utilizing AI in health care, clinicians, AI researchers, and also developers of AI devices and applications will be required to work together [5]. With the help of these AI technologies, decision-making and operational procedures can be enhanced. Physicians must take a proactive role in modifying clinical practices by using computational tools and contributing to the advancement of clinical AI [5].

AI tools have significant potential in addressing challenges in COPD. They can aid in developing consistent classification criteria by analyzing large datasets, identifying patterns, and enabling accurate and personalized diagnosis and treatment. AI can enhance understanding of COPD pathogenesis by analyzing complex biological data, leading to the identification of therapeutic targets and more effective interventions. AI-powered predictive models help identify individuals at high risk of exacerbations, enabling proactive management and cost reduction. By revolutionizing diagnosis, management, and treatment, AI tools offer personalized approaches, optimized resource allocation, and enhanced decision-making to improve COPD outcomes and reduce mortality rates.

III. ARTIFICIAL INTELLIGENCE BASED APPLICATION OF COPD

A. COPD detection and diagnosis

1) PULMONARY FUNCTION TESTS

Pulmonary function test (PFTs) is thought to be a crucial technique for assessing the respiratory system and for diagnosing disorders such as chronic obstructive pulmonary disease.

In one study, 50 people with varied pulmonary conditions were examined by 120 pulmonologists to determine PFTs patterns. Their analysis was contrasted with a decision tree-based machine learning method which uses general affected individual characteristics and functional parameters as input and the data was trained with 1,430 subjects [6]. The accuracy of pulmonary pattern interpretation among specialists were 74.4%, and the challenge of such a prominent task was reflected in the kappa interrater variability which is 0.67. In terms of diagnosing the appropriate PFT pattern, the automated technique showed 100% and 82% accuracy. These remarkable findings point to the potential of AI as a tool that supports decision making in clinical planning and proceedings.

It has been established that machine learning is extremely efficient in detecting patterns in feature spaces where a high number of dimensions or variables exist, and hence it can help with quicker and more precise diagnostic interpretation of PFT. Topalovic et al. [6] created a ML system that has a general accuracy of 68% for detecting the four most prevalent obstructive diseases: asthma, COPD, neuromuscular disease (NMD), and interstitial lung disease (ILD). The same study was used to demonstrate that when using the usual manual method of interpretation, only 38% of the candidates received the correct diagnosis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:44 UTC from IEEE Xplore. Restrictions apply.

To sum up, automating PFT interpretation with ML has largely solved numerous previous constraints of expert systems. Although it appears promising, more validation is necessary before PFT interpretation software can be used in a clinical context.

2) COMPUTED TOMOGRAPHY

In a set of data, AI systems look for patterns that can be utilized to estimate a clinical result or identify obstructive phenotypes. Automatic pulmonary function test interpretation and differential diagnosis of obstructive lung disorders are both successfully accomplished using ML. CT plays an important role in many areas. It is a modern tool for the diagnosis and evaluation of airway conditions as well as the detection and characterization of emphysema using images.

Convolutional neural networks (CNN) have been among the most effective deep learning architectures in the area of computer vision. A CNN model is made from one input layer, one output layer, and a set of many hidden layers. Every subsequent hidden layer's job is to perform a more detailed computation of the input data to get a better representation with increasing complexity, which is relevant to the categorization's objective [3]. CNN has been employed in the field of obstructive lung disorders to categorize textured patterns in CT images into several categories of emphysema tissue [7]. In an intriguing study, Tajbakhsh et al. [8] used a well-trained CNN model with exact tuning (transfer learning) to diagnose pulmonary embolism.

Positive outcomes have been obtained using CNN to diagnose obstructive patterns in CT scans. But creating a CNN from the start has its own challenges. It needs extensive computational resources, and huge training datasets—which are usually difficult to come by in the medical field—and is frequently affected by overfitting and convergence problems. A promising alternative is the use of a well-trained CNN network model on non-medical datasets of images in combination with the exact tuning of domain-related datasets [9].

B. Artificial intelligence based categorization of COPD

Depending upon indicatory symptoms, increasing exacerbation, and most importantly, the hospitalization history of COPD patients, they are categorized into four phenotypes justified by Global Obstructive Lung Disease Initiative. However, the unjust capability of this procedure is inadequate, which makes an AI/ML-type alliance of extra information which are the results of lung function tests, genome, physiological features, biomarkers, and comorbidities, for accurate categorization of phenotypes, severity evaluation, and therapeutic counseling more useful. In a study, an ML clustering model, named the k-means clustering has already been applied which helps in grouping and categorization of COPD patients into four phenotypes as different clusters. The method was performed on a dataset of 195 patients with medical history, physiological features, postbronchodilator FEV1, and COPD assessment test scores. FEV1 stands for how much air a person can forcefully expel from their lungs in just one second. The spirometer records the volume of air expelled during the first second of exhalation, which represents the FEV1 value. These values help doctors to understand the illness's severity and the development of the condition over time. When referring to the four clusters which are: Cluster (CL) 1, CL 2, CL3, and CL4 are termed as putative asthma overlap, mild, moderate, and severe COPD respectively, all of these clusters help to make different observations, such as Cluster 1 had the higher rate of severe exacerbation, whereas patients from Cluster 4 has forced expiratory volume in one second (FEV1) score of 46.7% of post-bronchodilator, high (17.5) COPD assessment test scores. [10]. However, the output needs to hold up for a longer period (>6 months). Another study performed on COPD patients from different Asian countries compared the data of change in FEV1, modified scale score of Medical Research Council dyspnoea, and St George's Respiratory Questionnaire (SGRQ) score for one year. The study reveals three phenotypes of COPD using hierarchical cluster analysis and factor analysis. Cluster 2 shows a high risk of FEV1 and acute exacerbation, whereas Cluster 1 and Cluster 3 show fever symptoms. Both studies have their own limitation one of them being less percentage of female participants (10%) into consideration. Also, the two clustering techniques (hierarchical clustering and k-means) and based on risk factors and co-occurring conditions, five phenotypes representing a large percentage of COPD patients were identified: weakness, cardiovascular disease, severe hindrance in airflow, anxiety/depression, diabetes, and non-comorbidity and obesity/atopy [11]. Even though the algorithms used in the above-mentioned studies are distinct and have certain limitations, they all reinforce the aim that inspecting various phenotypic categorizations can upgrade individualized diagnosis and medication. For example, if an individual is showing the symptoms of hindrance in airflow, then according to the AI tool the individual will be classified under Cluster 2 which shows that they have a severe condition of COPD which means that it requires immediate medication treatment.

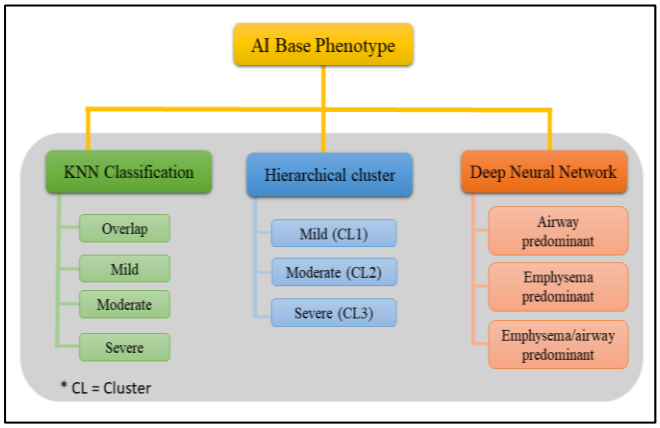


FIG. 1. Phenotype categorization using AI.

For a better understanding of phenotypes and accuracy, a Deep Neural Network model was developed. This model was trained on the 8980 individuals' dataset (COPDGene cohort study) and categorized phenotypes as normal, mixed emphysema/airway, emphysema predominant, and airway predominant. The methods demonstrated high precision of categorization for emphysema/small airway predominant phenotypes with an area under the receiver operating

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:44 UTC from IEEE Xplore. Restrictions apply.

characteristic curve (AUC) of 0.91 and emphysema/airway predominant phenotypes with an AUC of 0.80. During the comparison of the data with FEV1 prediction using random forest classification a large amount of data was found missing, since this ML technique was dependent on lung function tests for checking the hindrance of airflow and just a few patients are able to reach the end of the test, so it became a major drawback which needs to be corrected for advances in AI and ML.

Keeping the above problem in mind, an AI tool was developed that showed high accuracy of around 95% to find FEV1 values of COPD veterans (41,689) in the electronic health records (EHRs) [12]. The high accuracy is achieved as the training dataset that was used to train the model matched the model definition. So that's why when the training was done on the model it was able to achieve the desired accuracy within the margin of 70%-90%. Hence the model provided the required accuracy when used on testing data which has a similar categorization as in the training dataset.

Another study using logistic regression and support vector machine (SVM) algorithm-based ML model was used for analyzing the chest CT to assess the pulmonary ventilation function of COPD patients. The method was based on 87 features of an image that was taken and tested on 27 COPD patients; the model gave 87% of accuracy with 0.81 of AUC and validated the patient's COPD condition.[13] The output obtained from the model was satisfying but the dataset was not recorded on a large scale thus the COPD range of patients on which the data were obtained was from moderate to severe, so COPD patients having a condign mild range also need to be included for better analysis. With the advancement of AL/ML a new study was proposed using CNN that is done to observe the chest-CT. The data was collected from the group of regular smokers having COPD, which came from ECLIPSE cohorts and COPDGene to access the mortality, stage, diagnosis, and exacerbation. The data were divided into two phases, one being the training phase with 8,983 COPD smoker patients and another one being the testing phase with 1,672 patients. The output from the testing phase gave a 0.856 c-index value for the identification of COPD with 51.2% of precision for exactly determining the COPD patient's current stage. Moreover, the observations obtained for determining mortality and exacerbation of c-indices are 0.60 and 0.55 for the ECLIPSE cohort as well as 0.72 and 0.64 for the COPDGene cohort [14]. These outcomes show that the CNN model works better on the COPDGene cohort.

The approach used in the study has limited applicability because of its immense training time as well as computational resources. It becomes effortless for evaluating the seriousness of the data that can be obtained from clinical practice. In a study, clinical practice data of hospitalized acute exacerbations of COPD (AECOPD) were used to train an altered decision tree model that analyzes 28 clinical features. The data were taken from 208 patients having mild AECOPD and 202 patients having severe AECOPD. The accuracy obtained from testing the model was about 80.3% which was then used to identify the stage of AECOPD in hospitalized patients [15]. However, in the future other features can also be added to the dataset to improve the scope.

C. Supervision and surveillance of COPD

Readmissions and acute exacerbations may return as a result of airflow restriction and continuous chronic airway sensitivity of COPD. Multiple researchers adapted to new strategies based on ML, that have been shown to manage COPD patients and track the condition more effectively than traditional methods. In particular, 44,929 COPD hospitalizations split into test (30%) and training (70%) sets were examined using machine learning algorithms like deep neural networks and lasso regression. The built models attempted to identify if the patient would be readmitted within the period of 30 days of release in order to demonstrate the better capacity to predict (c-statistic = 0.61) as compared to the conventional techniques. In a similar manner, the medical claims data from COPD patients was mined using different non-deep and deep Machine learning algorithms in an attempt of predicting whether the readmission is happening within thirty days after release. The value 0.653 was found for the optimum AUC for the data depicting the high accuracy of predictions[16]. Decision tree analysis was used in a retrospective study of France to forecast the readmission of 143,006 COPD individuals over the age of 40. The study not only showed that the number of prior admissions was the most significant risk factor for readmission, but it also calculated the expense of readmission in 6 months. The prediction models can be utilized as a guide by physicians, despite the fact that these studies have a number of drawbacks, such as the absence of crucial clinical parameters.

COPD is a chronic, potentially fatal condition marked by a persistent restriction of airflow and ongoing respiratory symptoms. Thus, it is crucial to keep an eye on changes for lung function and avoid long-term airflow restriction. In order to identify those who are most likely to acquire or already have COPD, a machine learning model using random forest has been created utilizing spirometry data collected from 4,167 individuals. FEV1 was the model's primary output, and the threat of airflow limitation (FEV1/forced vital capacity) was its secondary output. This model might be helpful for predicting an individual's risk of hindrance in airflow and for preventing COPD in the initial stage [17].

Due to COPD's irreparability, early diagnosis and identification are essential, Thus, to demonstrate that six models using ML techniques were developed that use the data from the different clinical features and 101 single nucleotide polymorphisms. From the observation, it was found that nine single nucleotide polymorphisms have been shown to be strongly linked to this disorder. The ML models such as logistic regression and K-nearest neighbor classification tend to have a better level of accuracy and precision (81% and 82%, respectively) in the test set, while the algorithm used for building multilayer perceptron classification model is artificial neural network has achieved the maximum sensitivity of 100%. This model is perhaps useful for diagnosing COPD in the early stage despite only including a small number of genes and clinical variables, making up for test results of lung function tests for detection of the disease at its initial stage.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:44 UTC from IEEE Xplore. Restrictions apply.

For the early detection of AECOPD patients and the prediction of exacerbation, ML algorithms were utilized to assess functional respiratory imaging [18]. Similar to this, data from 135 AECOPD and 168 control participants' electronic health records were mined using various machine learning methods, such as SVM, k-nearest neighbor, random forest, logistic regression, and naive Bayesian. Additional testing and evaluation for the created models revealed that the SVM approach performed better with an AUC value of 0.90. As a result, ML models particularly the SVM model can aid medical professionals in locating AECOPD patients and making quick conclusions. However, the effectiveness of the models needs to be additionally verified using data from outside sources. Another strategy involved developing a mobile telemedicine system to enhance COPD self-management and quickly identify acute exacerbations in stable COPD patients. The technology may continuously track the clinical data of the registered patients in residence and issue a three-day advance warning for an acute exacerbation[19]. Even though the precision was just 40% and the study only lasted six months, it is still important to look at the advancement of an effective AI driven monitoring and warning mechanism. Every year, the cost of COPD on the world economy rises; this tendency is made worse by the aging of the population. According to estimates, COPD accounts for 56% of yearly healthcare expenses for respiratory disorders in Europe. Three ML algorithms - extreme gradient boosting, random forest, and logistic regression have been employed to evaluate 54 distinct demographic variables and clinical records from 780,295 hospitalizations in order to support vital information regarding health management and to determine and estimate the COPD patient cost in China. Though all ML models demonstrated outstanding predictive effectiveness, the extreme gradient boosting model had the best precision of 71.30% and an AUC value of 0.80, suggesting it could be useful for the predictive strategy for physicians, patients, and other healthcare specialists in progressive nations. Some studies have employed AI technology to determine the probability of COPD patient death because it's one of the most common causes of death in the world. As an example, the data of 3,9000 patients were collected who have COPD ranging from moderate to severe in order to predict death using the random forest classification model the data was observed over chest imaging characteristics and lung function tests. The 6-minute walk test, the ratio of pulmonary artery-to-aorta, and the FEV1 value were the best risk predictors according to the novel model, which demonstrated good predictive accuracy (C-index > 0.7) [20]. As a result, adapting to a new ML model that can detect COPD at its early stage can be helpful to prevent future damage, however, larger exterior population cohorts are required for the validation of the same.

D. Treatment of COPD

AI systems can track, integrate, and evaluate a vast range of heterogeneous clinical data from COPD patients, recommend the best course of action for each patient, and minimize under or over-treatment, brought on by mistakes made by clinicians. In a study, a variety of ML models were created using patient data of 135 individuals having COPD ranging from mild to extreme condition, the data was recorded on 153 prognostic parameters obtained from symptoms, baseline data, and telemonitoring of physiological data. These models include multitask neural network models, ensembles of boosted classifiers, and sparse maximum margin classifiers. The above-mentioned data was collected with the intention of identifying acute exacerbations and directing COPD corticosteroid medication. The best ML model (multitask neural network) demonstrated a greater value of AUC compared to non-ML approaches regardless of the use of chronic exacerbations or corticosteroids (0.73-0.76 vs 0.61-0.67) [21]. However, the model accuracy was measured using cross-validation rather than numerous independent cohorts, indicating the need for additional research.

IV. FUTURE POTENTIAL

Although, there has been significant progress in many areas of managing and understanding COPD, still a lot to be done in the future. The field of personalized medication for COPD is still developing as researchers look for distinctive phenotypes, biomarkers, prognostic indices, etc. that should direct a patient's decision-making process. AI can be used to make this procedure easier. AI can produce models using enormous volumes of data that are not readable to a doctor. The next phases in managing COPD are integrated computer systems with clinically related patient results, multi-scale and specific medical problems such as estimations of mortality, exacerbation risk, screening, and modifying medication depending on risk factors or characteristics of the disease.

AI has many uses one of them is its potential to define the mathematical relationship between real-life data sets. The current way of identification for the stage of COPD phenotype helps in clinical practices. However, all these methods are hypothetical and to solve this issue the image based phenotype can be used in combination with molecular and genetic features that allow us to explore the disease which in turn leads us to new remedies that are less hypothetical [22]. Even though the results obtained from the AI integration can be speculative, they can still be used for the understanding and uncovering of different areas of disease nature. The integration of AI with clinical data has changed perspectives about many diseases and discovered new potentials in the field of diagnosis and treatment of COPD.

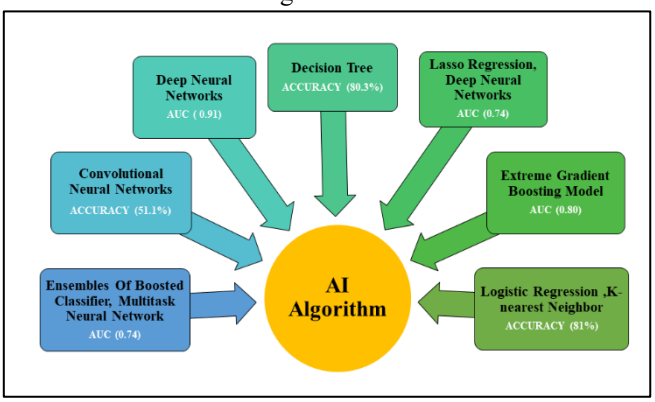


FIG. 2. Machine Learning models with their respective measures.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:44 UTC from IEEE Xplore. Restrictions apply.

V. CONCLUSION

Artificial intelligence has gained widespread acceptance in healthcare fields and has begun to significantly dominate the field of respiratory research. AI includes machine learning, which allows computers to learn on their own without explicit programming. This technology is changing the clinical perception regarding imaging and diagnostic data and their use to investigate the causal factors for the progression of severe diseases such as COPD. The grasp of disease trajectories and subtypes can be expanded beyond the existing knowledge of disease's phenotype with the help of a machine learning model which combines biomarker, functional, imaging, and multi-omics data. The paper revolved around different types of data sets and each of them has different nature. Some data have a high male to female ratio others have different levels of symptoms. Each data was trained using different models technique and provided different accuracy results which provide better Interpretability for our datasets hence the model with high accuracy and better Interpretability was pricked through the course of this paper, we discovered and uncovered many areas of AI application in COPD, defined tools and studies for characterization, assessment, and monitoring practices, as well as therapeutic practices.

ACKNOWLEDGMENT

The work was supported and guided by the members of the Department of Biotechnology, Delhi Technological University and hence I am very thankful to them.

REFERENCES

[1] K. Matsunaga, M. Harada, J. Suizu, K. Oishi, M. Asami-Noyama, and T. Hirano, "Comorbid Conditions in Chronic Obstructive Pulmonary Disease: Potential Therapeutic Targets for Unmet Needs," *J Clin Med*, vol. 9, no. 10, p. 3078, Sep. 2020, doi: 10.3390/jcm9103078.

[2] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.

[3] A. Sheikhtaheri, F. Sadoughi, and Z. Hashemi Dehaghi, "Developing and Using Expert Systems and Neural Networks in Medicine: A Review on Benefits and Challenges," *J Med Syst*, vol. 38, no. 9, p. 110, Sep. 2014, doi: 10.1007/s10916-014-0110-5.

[4] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/NEJMra1814259.

[5] W. W. Stead, "Clinical Implications and Challenges of Artificial Intelligence and Deep Learning," *JAMA*, vol. 320, no. 11, p. 1107, Sep. 2018, doi: 10.1001/jama.2018.11029.

[6] M. Topalovic, S. Laval, J.-M. Aerts, T. Troosters, M. Decramer, and W. Janssens, "Automated Interpretation of Pulmonary Function Tests in Adults with Respiratory Complaints," *Respiration*, vol. 93, no. 3, pp. 170–178, 2017, doi: 10.1159/000454956.

[7] D. Bermejo-Pelaez, R. S. J. Estepar, and M. J. Ledesma-Carbayo, "Emphysema classification using a multi-view convolutional network," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, Apr. 2018, pp. 519–522. doi: 10.1109/ISBI.2018.8363629.

[8] N. Tajbakhsh *et al.*, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016, doi: 10.1109/TMI.2016.2535302.

[9] H.-C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.

[10] H.-Y. Yoon *et al.*, " Prediction of first acute exacerbation using COPD subtypes identified by cluster analysis ," *Int J Chron Obstruct Pulmon Dis*, vol. Volume 14, pp. 1389–1397, Jun. 2019, doi: 10.2147/COPD.S205517.

[11] M. Pikoula, J. K. Quint, F. Nissen, H. Hemingway, L. Smeeth, and S. Denaxas, "Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records," *BMC Med Inform Decis Mak*, vol. 19, no. 1, p. 86, Dec. 2019, doi: 10.1186/s12911-019-0805-0.

[12] K. M. Akgün *et al.*, "Extracting lung function measurements to enhance phenotyping of chronic obstructive pulmonary disease (COPD) in an electronic health record using automated tools," *PLoS One*, vol. 15, no. 1, p. e0227730, Jan. 2020, doi: 10.1371/journal.pone.0227730.

[13] A. Westcott, D. P. I. Capaldi, D. G. McCormack, A. D. Ward, A. Fenster, and G. Parraga, "Chronic Obstructive Pulmonary Disease: Thoracic CT Texture Analysis and Machine Learning to Predict Pulmonary Ventilation," *Radiology*, vol. 293, no. 3, pp. 676–684, Dec. 2019, doi: 10.1148/radiol.2019190450.

[14] G. González *et al.*, "Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography," *Am J Respir Crit Care Med*, vol. 197, no. 2, pp. 193–203, Jan. 2018, doi: 10.1164/rccm.201705-0860OC.

[15] J. Peng, C. Chen, M. Zhou, X. Xie, Y. Zhou, and C.-H. Luo, "A Machine-learning Approach to Forecast Aggravation Risk in Patients with Acute Exacerbation of Chronic Obstructive Pulmonary Disease with Clinical Indicators," *Sci Rep*, vol. 10, no. 1, p. 3118, Feb. 2020, doi: 10.1038/s41598-020-60042-1.

[16] X. Min, B. Yu, and F. Wang, "Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on COPD," *Sci Rep*, vol. 9, no. 1, p. 2362, Feb. 2019, doi: 10.1038/s41598-019-39071-y.

[17] W. Chen, D. D. Sin, J. M. FitzGerald, A. Safari, A. Adibi, and M. Sadatsafavi, "An Individualized Prediction Model for Long-term Lung Function Trajectory and Risk of COPD in the General Population," *Chest*, vol. 157, no. 3, pp. 547–557, Mar. 2020, doi: 10.1016/j.chest.2019.09.003.

[18] M. Lanclus *et al.*, "Machine Learning Algorithms Utilizing Functional Respiratory Imaging May Predict COPD Exacerbations," *Acad Radiol*, vol. 26, no. 9, pp. 1191–1199, Sep. 2019, doi: 10.1016/j.acra.2018.10.022.

[19] M. Hardinge *et al.*, "Using a mobile health application to support self-management in chronic obstructive pulmonary disease: a six-month cohort study," *BMC Med Inform Decis Mak*, vol. 15, no. 1, p. 46, Dec. 2015, doi: 10.1186/s12911-015-0171-5.

[20] M. Moll *et al.*, "Machine Learning and Prediction of All-Cause Mortality in COPD," *Chest*, vol. 158, no. 3, pp. 952–964, Sep. 2020, doi: 10.1016/j.chest.2020.02.079.

[21] P. Orchard *et al.*, "Improving Prediction of Risk of Hospital Admission in Chronic Obstructive Pulmonary Disease: Application of Machine Learning to Telemonitoring Data," *J Med Internet Res*, vol. 20, no. 9, p. e263, Sep. 2018, doi: 10.2196/jmir.9227.

[22] A. Agustí, B. Celli, and R. Faner, "What does endotyping mean for treatment in chronic obstructive pulmonary disease?," *The Lancet*, vol. 390, no. 10098, pp. 980–987, Sep. 2017, doi: 10.1016/S0140-6736(17)32136-0.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:44 UTC from IEEE Xplore. Restrictions apply.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Apple Leaves Disease Detection Using Multilayer Convolutional Neural Network

Saradindu Das
Department of Software Engineering
Delhi Technological University
Delhi, India
saradindu24@gmail.com

Abhilasha Sharma
Department of Software Engineering
Delhi Technological University
Delhi, India
abhi16.sharma@gmail.com

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205764

***Abstract*—Agriculture is the process of growing crops and raising livestock and cultivating other forms of food. It has been a fundamental activity for human societies throughout history, providing food and other resources necessary for survival. Hence plant diseases can have a significant impact on the economy, particularly in agricultural-dependent countries. Crop yield loss, trade restrictions, increased production costs and reduced agricultural productivity are some of the ways plant diseases can affect the economy. Farmers visually inspect their crops for symptoms of diseases, such as discoloration, spots, wilting, and deformities. Farmers can also use their sense of touch and smell to detect diseases, such as the sticky or slimy feel of plant leaves infected with fungal diseases and the foul smell of rotting or decaying plant material. However, traditional methods of plant disease detection do have limitations. Visual inspection and other traditional methods may not always detect diseases at an early stage, and there is a risk of misdiagnosis. Furthermore, conventional techniques may fail to identify illnesses that are not perceptible to the unaided observation. The other alternatives are the use of Artificial Intelligence(AI) which includes training computers to detect plant diseases using image recognition technology. These AI methods are increasingly being used for plant disease detection because they offer a fast, accurate, and cost-effective way to diagnose plant diseases, which can help prevent crop losses and increase yields. In this article, a Multilayer CNN model, inspired by InceptionNet, is put forward. This model is trained on the "Plant Pathology 2021: FGVC8 dataset". This proposed model is compared with three existing CNN models using RMSprop with Nesterov Momentum Optimizer. The suggested model outperforms other models that are already trained and achieves an accuracy of 92.56% in diagnosing the disease.**

***Index Terms*—Artificial Intelligence (AI), Deep Learning (DL), Convolutional Neural Network (CNN), Plant Disease**

I. Introduction

People all over the world enjoy eating apples, which are among the most common and beloved fruits. They are not only delicious but also packed with many essential nutrients and health benefits. Here are some of the benefits of apples: Nutrient-rich, Good for heart health, boosts immune system, may help regulate blood sugar levels. In general, apples are a flavourful and nutritious fruit that can offer numerous health advantages if eaten as a component of a well-rounded diet. However, there are several diseases that can affect apple plants, some of which are caused by fungi, bacteria, or viruses. These diseases[1] in the plants can be harmful in several ways, both to individual plants and to entire ecosystems. Plant diseases[2] can have significant economic impacts, both in terms of lost productivity for farmers and increased costs for consumers. It can have adverse effects on ecosystem function: and can affect the interactions between plants and other organisms, such as pollinators or herbivores, leading to changes in ecosystem function.

Detecting plant diseases[2] in their early stages is crucial for effective management and control of the disease. Managing apple diseases typically involves a combination of cultural, biological, and chemical control methods, including proper pruning and sanitation, use of disease-resistant varieties, and application of fungicides and bactericides when necessary. Traditionally, plant disease[2] detection involves visual inspection of plants by experienced plant pathologists or farmers. The presence of characteristic symptoms, such as spots, wilting, or deformities, is often used to diagnose the disease. In some cases, laboratory tests may be required to confirm the diagnosis. While these traditional methods are still commonly used for plant disease[2] detection, they can be time-consuming, expensive, and require specialised expertise.

In recent years, new technologies such as remote sensing, AI[3], and biosensors have emerged as promising tools for early and rapid detection of plant diseases[2]. These new approaches offer faster, more accurate, and cost-effective solutions for disease detection and management. The next alternatives are Imaging techniques which have emerged as promising tools for prior detection of plant diseases[2]. They have the potential to revolutionize plant disease[2] detection by offering fast, non-destructive, and cost-effective solutions for prior detection and management of plant diseases. While imaging techniques offer several advantages for plant disease[2] detection, there are also some disadvantages to consider: Imaging equipment can be expensive, and specialized training may be required to use and interpret the images. Imaging techniques may only capture images of small areas of plant tissue, requiring multiple images to be taken to cover larger plants or fields. Environmental elements like lighting or humidity can impact the image quality produced by imaging technologies. ML[4] techniques like Support Vector Machines(SVM)[5] and K-Means Clustering[6] have shown promise in the detection and diagnosis of plant diseases[2]. But it comes with the added

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:04 UTC from IEEE Xplore. Restrictions apply.

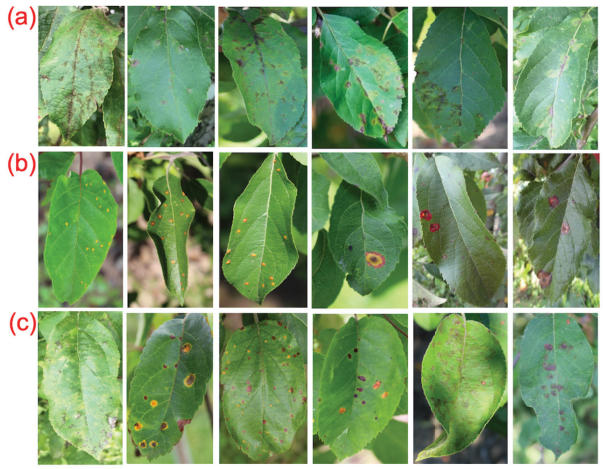


Fig. 1. Images of disease symptoms on leaves[7]
(a) A leaf illuminated by sunlight that is not direct.
(b) A leaf illuminated by sunlight that is shining straight onto it.
(c) A leaf with a prominent reflection.

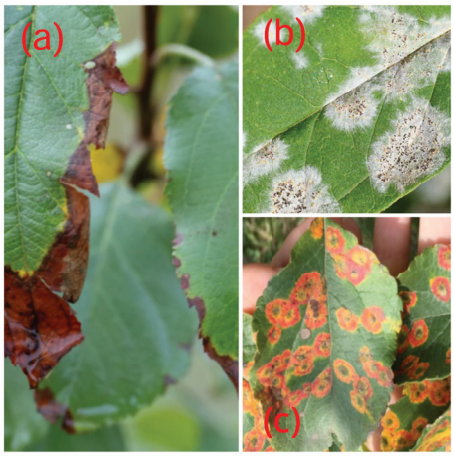


Fig. 2. Images of different apple leaf diseases[8]–[10]
(a) Fire blight
(b) Powdery mildew on leaf
(c) Cedar apple rust infection on the leaf

disadvantage of a complex feature extraction process which reduces the overall efficiency of the system.

On the other hand, we have DL[11] algorithms such as recurrent neural networks (RNNs)[12], CNNs etc. These algorithms are trained on large datasets of plant images to learn the characteristic features of healthy and diseased plants, enabling automated disease detection and classification with high accuracy. DL[11] algorithms can be trained on large datasets of plant images to automatically detect disease[2] symptoms with high accuracy. This can enable faster and more accurate disease detection, particularly in large-scale agricultural settings.

DL[11] algorithms can detect subtle changes in plant tissue that may be difficult for humans or traditional ML[4] algorithms to detect, leading to improved accuracy in disease detection[2]. Overall, We can use DL[11] methods to find and deal with plant diseases[2] before they get worse, with the potential to improve crop yields, reduce pesticide use, and promote sustainable agriculture. While DL[11] techniques offer many advantages for plant disease[2] detection, there are also several challenges that need to be addressed. Plants can vary widely in their morphology and appearance, making it difficult to develop generalized DL[11] models that can detect a wide range of diseases across different plant species. Labelling and annotating large datasets of plant images can be time-consuming and resource-intensive, particularly for rare or emerging plant diseases.

Here, in this study the FGVC8 dataset from the Plant Pathology 2021 Kaggle competition[13], is taken. The FGVC8[13] dataset is a collection of images used for fine-grained visual categorization tasks, which includes the classification and recognition of subordinate categories within a larger category. The apple leaves with unhealthy areas are shown in Fig. 1 and in Fig. 2 which shows the various unhealthy leaves of apple trees. Multilayer CNN model makes it quite simple to identify them. The proposed Multilayer CNN model will be used to identify such apple foliar diseases[1], and the performance will be compared to previously trained models which are based on InceptionNet[14].

Taking inspiration from InceptionNet[14], the proposed Multilayer CNN model is comprised of a Convolution Layer, Batch Layer, Max-Pooling Layer, Flatten Layer, Dense Layer, and Inception Block. The result shows that the proposed model, which is the Multilayer CNN model works very well with RMS Prop with Nesterov momentum Optimizer[14] which is trained on Plant Pathology 2021- FGCV8 Dataset[13]

Fundamentals will be covered in chapter 2. Related work will then be covered in chapter 3 along with the methods that have previously been suggested for identifying PDs. chapter 4 will provide a summary of the proposed work. Data evaluation will be described in chapter 5. Experiments and results will be discussed in chapter 6, followed by the conclusion and future scope in chapter 7 and references in the last section.

II. Background

When identifying a foliar disease[1] in apple leaves from an image, one DL[11] algorithm known as CNNs performs better than other deep learning algorithms, resulting in higher accuracy. A CNN is a sort of deep neural network that is often utilized in computer vision applications such as image and video identification. It is an advanced Neural Network[15]-[17] architecture made specifically or processing data with a grid-like layout, like pictures, by performing a series of convolutional and pooling operations. Convolutional layers, pooling layers, and fully connected layers are the key building blocks of a CNN. By applying different filters to the input image, these layers enhance the accuracy of the model. It can have varying numbers of layers depending on the architecture and complexity of the model. The different layers are the following.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:04 UTC from IEEE Xplore. Restrictions apply.

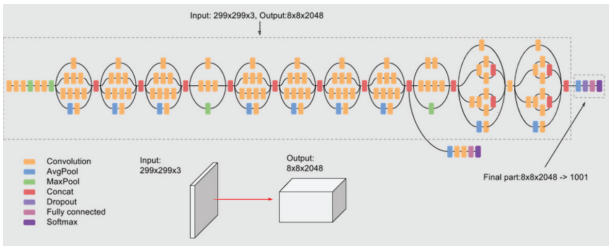


Fig. 3. Architecture of InceptioNet[20]

1) Convolutional layers: The convolutional layer is a key building block in CNNs. It performs a mathematical operation called convolution.
2) Pooling layers: In CNNs, the pooling layer serves as a crucial important block, contributing to dimensionality reduction and feature extraction. Its main function is to reduce the spatial dimensions of the feature maps produced by convolutional layers while retaining the most important feature.
3) Connected layers: The connected layer, also can be called as the dense layer, is a key component in CNNs. It is typically the last layer in the network and is responsible for making the final prediction by mapping the extracted features to a specific class or label.

The filters start as simple features in the lowest layer and become more complex as the layers progress to identify unique features of the input object. In the classification of plant diseases[2], leaf images are utilized because they exhibit signs sooner than rest of the plant, making it easier to detect and diagnose the disease[2] at an early stage. CNNs use features and texture analysis to accomplish this task. The fully connected layer, which uses the flattened vector from the preceding layers as its input, computes the output vector for the CNN. The fully connected layer picks up a set of weights and biases that translate the input vector to the predicted probabilities for each class in the output vector. To enhance the CNN's performance for a given job, the fully connected layer's neuron count can be altered.

Google created a type of neural network[15]–[17] called the InceptionNet[18] for images that applies different types and sizes of filters simultaneously, and merges their results to produce a single result. This helps the network to learn features at different levels of detail and resolution, and makes it more accurate in tasks that involve images. The network has been applied to various problems, such as identifying images, finding objects, and segmenting images. The first version of the network, also called GoogLeNet[19], had 22 layers, including layers for filtering, pooling, and connecting. However, there have been later versions and enhancements of the network that have more layers, such as InceptionV2, InceptionV3, and InceptionV4. These newer versions have up to hundreds of layers and use additional techniques like batch normalization and residual connections to make them better. The architecture of InceptionNet[18] is given in Fig 3.

III. Related Work

In this chapter, we will look at a few of the important researches by other authors. Several DL[11] techniques have been introduced in the past few years.

In the paper, proposed by[21] the DL approach is used which is based on CNN. An available dataset, from PlantVillage[22] with 39 different classes is utilised to train the suggested CNN model. Before training, this dataset undergoes 6 different types of data augmentation methods. This proposed model is compared with transfer learning approaches. It has been concluded that compared to popular transfer learning[23], [24] approaches, data augmentation can improve model performance. With transfer learning, DL[11] models can be built with less training data, less time, and fewer computational costs.

The author of [25] suggested a model that uses transfer learning, called INC-VGG, which merges VGG19[26] and InceptionV3. After the first layers of VGG19 models, three ConvBN layers of size 3X3 were added, followed by two InceptionV3 layers, and in the end a Softmax layer. The suggested model merges two models that are already trained and it outperforms other models that are also trained beforehand.

A multilayer model based on CNN was proposed by [27] to identify tomato diseases present on monitored tomato plants, compared CNN with F-RCNN[28], and used Transfer Learning[23],[24] for disease recognition on tomato plant leaves. Before training, data augmentation is done to increase the dataset size and reduce overfitting. Transfer Learning is used as well. In comparison to F-RCNN[28], the proposed model gives better accuracy.

Author [29] proposed the DenseNet201 model which contains 201 layers. This proposed model is compared with transfer learning approaches. As a result of this proposed model, the vanishing gradient problem has been solved, feature propagation has been improved, feature reuse has been promoted, and the number of parameters has been reduced. Networks with too many connections are less efficient in computation, and are more prone to overfitting.

Author of [30] specifies in his study a technique that has used a model called LeNet[31], which has layers for filtering, activating, pooling, and connecting, to train it on data and find diseases in tomatoes. This model can correctly recognize different kinds of diseases in tomato plants. The model gets the data after it has been augmented. It achieves about 95% of accuracy which is much higher than other models that use CNN or are based on CNN.

The researchers [32] suggested a DCNN that uses Bayesian learning[33] on a residual network. This method combines the strength of Bayesian learning[33] with the speed of residual networks for better performance in finding diseases. This model is trained to find diseases in tomatoes, potatoes, and bell peppers. The process of learning features is effective but it takes a lot of time to compute.

[34]Developed a Deep Siamese CNN. Photos of grape leaves were collected and sorted into four categories. Photographs are taken from the field.The proposed model handles

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:04 UTC from IEEE Xplore. Restrictions apply.

the challenge of a small dataset. Although it gives high accuracy, it is not computationally efficient.

[35] studied the AlexNet[36] and GoogLeNet[19] DNN architectures and achieved a classification accuracy of 99.55% on the PlantVillage[22] dataset using 50k sample images over 30 training epochs. Images captured from the farm. This model is prone to overfitting because it uses only 800 leaf images. And also a simple model which uses only three layers.

[37] developed a three-layer CNN system with an average accuracy of 94.9% after 40 epochs of training. The Images are captured from the farm. This study used 800 leaf images of cucumber. This model is prone to overfitting because it uses only 800 leaf images. And also a simple model which uses only three layers.

The article[38] examines the utilization of DL[11] algorithms to identify plant diseases[2] through images. It presents a summary of DL[11] models, datasets, and preprocessing techniques employed to train such models. The article emphasizes the significance of data augmentation, transfer learning[23], [24], and hyperparameter tuning to attain high accuracy in Plant Disease[2] Diagnosis. Additionally, it discusses the obstacles and prospects of using DL-based image analysis in Plant Disease[2] Detection. In general, the article presents a comprehensive summary of the utilization of DL[11] algorithms for image-based Plant Disease[2] Diagnosis. It emphasizes the potential for these algorithms to transform agricultural production by enabling the timely and precise identification of plant diseases[2].

IV. Proposed Methodology

This chapter focuses on the Multilayer CNN model which has been proposed. The first section describes the dataset used in the study, followed by an explanation of the Data Preprocessing in the subsequent section. Finally, the Proposed Model is discussed in the final part of this chapter. It can be seen how the model that has been suggested works in detail in Fig. 4.1.

A. Dataset

One real-world datasets, Plant Pathology 2021: FGCV8[13] has been used to identify a certain disease affecting apple plants. This is a summary of the datasets that have been used.

TABLE I
Distribution of Dataset

Total Images	Training Images	Testing Images
18632	16769	1863

B. Data Acquisition

Data collecting from various sources and conversion to a digital representation is known as data acquisition. It then can be handled by a computer or another electronic device. Overall, data acquisition plays a crucial role in collecting and analyzing data from various sources, allowing for better decision making, process control, and scientific research.

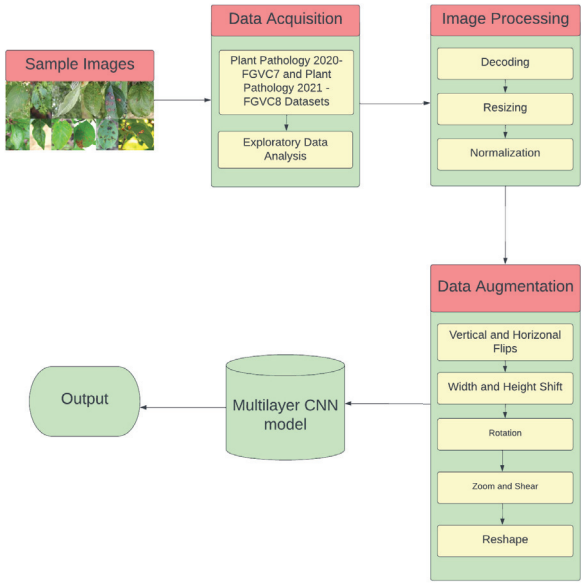


Fig. 4. Graph showing proposed methodology process

C. Image Processing

The editing and analysis of digital pictures fall under the umbrella of the field of image processing. It involves the use of various algorithms to transform, enhance, or extract information from images. Some of the common techniques used in image processing include image filtering, edge detection, image segmentation, image restoration, and feature extraction. These techniques can be used individually or in combination to achieve various image processing goals. In recent years, DL[11] techniques such as CNNs have also been used in image processing, achieving trailblazing results in various image associated tasks such segmentation, and image grouping. It uses various algorithms and techniques to change, improve, or get information from images. These techniques can be used separately or together to achieve different image processing goals. In recent years, DL[11] techniques such as CNNs have also been used in image processing, getting the best results in various tasks related to images such as object detection, segmentation, and image classification.

D. Data Augmentation

It is the process of making a dataset bigger by applying various changes or alterations to the original data. Data augmentation aims to broaden the dataset's diversity, which can make ML[4] models better by preventing overfitting and improving generalization. There are several techniques that can be used for data augmentation, based on the task and the type of data.

E. Proposed Model

A CNN based model called Multilayer CNN is proposed which contains Convolution Layer, Batch Normalization Layer, Max-Pooling Layer, Flatten Layer, Dense Layer,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:04 UTC from IEEE Xplore. Restrictions apply.

Dropout Layer and Inception Block. Here Inception Block is inspired from InceptionNet[14]. And in this proposed architecture, ReLU, LeakyReLU and Softmax activation functions are used. In Table 1, a multilayer CNN architecture is presented. Detailed Explanation of Each Layer is given below:

- Input: Defines the shape of the input tensor, which is (input image, input image, 3) for this model.
- Conv2D: Performs 2D convolution on the input tensor with a specified number of filters (64 in the first layer), kernel size of 3x3, and padding of 'same', which means the output feature maps will have the same spatial dimensions as the input.
- BatchNormalization: It ensures that the mean activation is near to 0 and the activation standard deviation is close to 1. Normalises the outputs of the preceding layer.
- Activation: Applies the ReLU activation function to the output of the previous layer.
- MaxPooling2D: Downsamples the input tensor along the spatial dimensions by taking the maximum value within a specified pool size (default is 2x2) and strides (default is pool size).
- inception_module: A custom layer that applies the inception module that takes the output of the previous layer as input and returns a feature map with increased depth through concatenation of different convolutional filters. inception_module consists of a 1x1 convolution branch, a 3x3 convolution branch, a max pooling, and a 1x1 convolution followed by 5x5 convolution and then 2x2 convolution branch. The output of each branch is concatenated along the channel axis to produce the final output.
- Flatten: Used to transform the 3D output of a convolutional layer into a 1D vector that can be given to a fully connected layer. The Flatten layer basically "flattens" the output of the preceding layer into a long, one-dimensional array of values.
- GlobalAveragePooling2D: Takes the avg. of each feature map in the preceding layer along the spatial dimensions to reduce the tensor to a vector of length equal to the number of filters.
- Dense: A fully connected layer that takes the vector output of the previous layer and applies a specified number of neurons (9216, 4096, 1024, and 128 in this model) with ReLU activation.
- Dropout: Randomly sets a fraction of input units to 0 at each update during training to prevent overfitting. BatchNormalization: Applies batch normalization to the output of the previous layer.
- Dense: A fully connected layer that takes the output of the previous layer and applies a softmax activation to classify the input into one of 4 possible classes(in case of Plant Pathology 2020-FGCV7) or 12 possible classes(in case of Plant Pathology 2021-FGCV8) [as defined by the number of neurons in the output layer].

S. No.	Name	Activations	Total Learnables
1.	Image_Input	256x256x3	0
2.	conv2d	256x256x64	1792
3.	batch_normalization	256x256x64	265
4.	activation	256x256x64	0
5.	max_pooling2d	128x128x64	0
6.	inception_module	128x128x16	0
7.	dense_module	128x128x54	0
8.	conv2d_1	128x128x32	156704
9.	batch_normalization_1	128x128x32	128
10.	activation_1	128x128x32	0
11.	max_pooling2d_1	64x64x32	0
12.	inception_module_1	64x64x80	0
13.	dense_module_1	64x64x272	0
14.	global_average_pooling2d	272	0
15.	dense	9216	2515968
16.	dense_1	4096	37752832
17.	dense_2	1024	4195328
18.	dense_3	128	131200
19.	dropout	128	0
20.	batch_normalization_3	128	512
21.	dense_4	4	1548
	Total params:		46,255,636

TABLE II
DETAILED ARCHITECTURE OF MULTILAYER CNN MODEL

F. Optimizer Used

In this suggested work, RMS Prop with Nesterov momentum optimizers[14] is used to test whether the suggested multilayer CNN model works better than the existing models. In order to decrease the loss function, method called RMS Prop Nesterov momentum[14] in DL[11] are used to modify the biases and weights of neural networks during training. The choice of optimizer can have a big impact on the performance of the model, and there are several different optimizers available to pick from.

V. EXPERIMENTAL EVALUATION

This chapter will cover the performance of the proposed model. In this case, the suggested model is evaluated using data from dataset. The Data Evaluation section contains information on these datasets in depth. In this study, the suggested model was trained using the optimizer, and the outcomes were compared. The optimizer has a learning rate[39] setting of 0.01 for learning. further contrasted the outcomes with three previously suggested CNN based algorithms.

A. Performance Matrics

- Accuracy: The degree to which a measurement, computation, or forecast is accurate or precise is referred to as accuracy. To express how precise a prediction or measurement is, we can use a fraction or a percentage.

$$Accuracy = \frac{numberofcorrectpredictions}{totalnumberofpredictions} \quad (1)$$

- Precision: Precision is the percentage of all retrieved occurrences that are relevant.

$$Precision = \frac{NumberofPredictedPositive}{NumberofTotalPositivePredictions} \quad (2)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:04 UTC from IEEE Xplore. Restrictions apply.

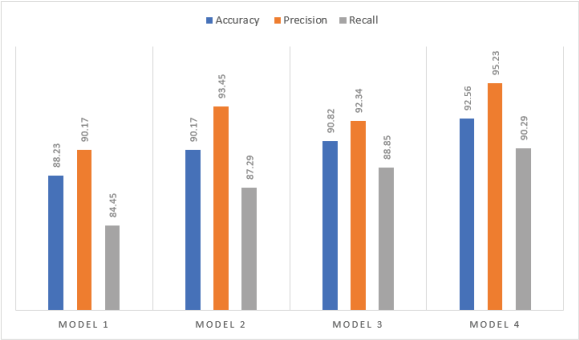


Fig. 5. Comparison of Performance[21], [40], [41]

- Recall: It is the ratio of true positives i.e., positively identified cases to the total of true positives and false negatives i.e., positively recognised cases that were overlooked.

$$Recall = \frac{Truepositives}{Truepositives + Falsenegatives} \quad (3)$$

Accuracy, Precision, and Recall have been used as assessment measures to assess the performance of the proposed approach on the dataset. Additionally, the performance of the proposed model has been compared using the optimizer, namely RMSprop with Nesterov Momentum[14] with 3 existing models.

The model is trained on the dataset for 50 epochs using the RMSprop with Nesterov Momentum Optimizer[14]. The model with the best accuracy (0.9256), precision (0.9523), recall (0.9029), and least loss (0.2543) is this one.

A Bar chart has been shown in the Fig. 5 which compares the performance of the proposed model with the other existing CNN-based models[21], [40], [41] using the same optimizer[14].

VI. Conclusion

CNNs have already shown great promise in the field of plant disease detection and diagnosis. However, there are several potential areas for future development and application of CNNs in this field. Prior to usage, the entire dataset is divided into two subsets: the training dataset and the test dataset. The split is performed at a ratio of 18:2, which corresponds to 90% and 10% of the dataset, respectively. The proposed model has proven to give a better performance of 92.56% than the existing models hence the proposed model can be applied for better performance. Overall, the proposed model has the capability to reorganize the field of plant disease detection and management by providing accurate, efficient, and early detection of diseases, which can help in maintaining crop health and improving agricultural productivity.

References

[1] "Apple: Diseases and Symptoms — Vikaspedia." https://vikaspedia.in/agriculture/crop-production/integrated-pest-managment/ipm-for-fruit-crops/ipm-strategies-for-apple/apple-diseases-and-symptoms (accessed Apr. 06, 2023).

[2] "plant disease - Symptoms and signs — Britannica." https://www.britannica.com/science/plant-disease/Symptoms-and-signs (accessed May 25, 2022).

[3] J. Mccarthy, "WHAT IS ARTIFICIAL INTELLIGENCE?," 2007, Accessed: Mar. 31, 2023. [Online]. Available: http://www-formal.stanford.edu/jmc/

[4] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science (1979), vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/SCIENCE.AAA8415.

[5] D. A. Pisner and D. M. Schnyer, Support vector machine. Elsevier Inc., 2019. doi: 10.1016/B978-0-12-815739-8.00006-7.

[6] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," Appl Stat, vol. 28, no. 1, p. 100, 1979, doi: 10.2307/2346830.

[7] "[PDF] The Plant Pathology 2020 challenge dataset to classify foliar disease of apples — Semantic Scholar." https://www.semanticscholar.org/paper/The-Plant-Pathology-2020-challenge-dataset-to-of-Thapa-Snavely/17005a1bd4189707f17d8bef9a0909c9399f7171 (accessed Mar. 18, 2023).

[8] "Cedar Apple Rust — NC State Extension Publications." https://content.ces.ncsu.edu/cedar-apple-rusts (accessed Apr. 21, 2023).

[9] "Mildew is, luckily, an easy to fix disease — PLNTS.com." https://plnts.com/en/care/diseases/mildew (accessed Apr. 21, 2023).

[10] "Tips Information about Plant Diseases - Gardening Know How." https://www.gardeningknowhow.com/plant-problems/disease (accessed Apr. 21, 2023).

[11] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, May 27, 2015. doi: 10.1038/nature14539.

[12] L. Medsker, D. L. Jain, and B. Raton London New York Washington, "RECURRENT NEURAL NETWORKS E dited by Design and Applications," 2001.

[13] "Plant Pathology 2021 - FGVC8 — Kaggle." https://www.kaggle.com/c/plant-pathology-2021-fgvc8 (accessed Dec. 14, 2022).

[14] J. Zhong, C. Pun, and S. Member, "An End-to-End Dense-InceptionNet for Image Copy-Move Forgery Detection," vol. 15, pp. 2134–2146, 2020.

[15] S. Ruder, "An overview of gradient descent optimization algorithms," Sep. 2016, Accessed: Apr. 22, 2023. [Online]. Available: https://arxiv.org/abs/1609.04747v2

[16] C. M. Bishop, "Neural networks and their applications," Review of Scientific Instruments, vol. 65, no. 6, p. 1803, Jun. 1998, doi: 10.1063/1.1144830.

[17] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," Computer (Long Beach Calif), vol. 29, no. 3, pp. 31–44, Mar. 1996, doi: 10.1109/2.485891.

[18] "Artificial neural network - Wikipedia." https://en.wikipedia.org/wiki/Artificial_neural_network (accessed Apr. 06, 2023).

[19] "Deep Learning: GoogLeNet Explained — by Richmond Alake — Towards Data Science." https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765 (accessed May 26, 2022).

[20] "Inception-v3 Explained — Papers With Code." https://paperswithcode.com/method/inception-v3 (accessed Apr. 22, 2023).

[21] G. Geetharamani and A. P. J., "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," Computers and Electrical Engineering, vol. 76, pp. 323–338, Jun. 2019, doi: 10.1016/j.compeleceng.2019.04.011.

[22] "GitHub - spMohanty/PlantVillage-Dataset: Dataset of diseased plant leaf images and corresponding labels." https://github.com/spMohanty/PlantVillage-Dataset (accessed May 25, 2022).

[23] S. Panigrahi, A. Nanda, and T. Swarnkar, "A Survey on Transfer Learning," Smart Innovation, Systems and Technologies, vol. 194, no. 10, pp. 781–789, 2021, doi: 10.1007/978-981-15-5971-6_83.

[24] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10. pp. 1345–1359, 2010. doi: 10.1109/TKDE.2009.191.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:04 UTC from IEEE Xplore. Restrictions apply.

[25] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," Comput Electron Agric, vol. 173, p. 105393, Jun. 2020, doi: 10.1016/J.COMPAG.2020.105393.
[26] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," Proceedings of IEEE International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications, CENTCON 2021, pp. 96–99, 2021, doi: 10.1109/CENTCON52345.2021.9687944.
[27] R. G. De Luna, E. P. Dadios, and A. A. Bandala, "Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and Recognition," IEEE Region 10 Annual International Conference, Proceedings/TENCON, vol. 2018-October, pp. 1414–1419, Feb. 2019, doi: 10.1109/TENCON.2018.8650088.
[28] "Faster R-CNN Explained for Object Detection Tasks — Paperspace Blog." https://blog.paperspace.com/faster-r-cnn-explained-object-detection/ (accessed Apr. 06, 2023).
[29] V. Tiwari, R. C. Joshi, and M. K. Dutta, "Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images," Ecol Inform, vol. 63, no. March, p. 101289, 2021, doi: 10.1016/j.ecoinf.2021.101289.
[30] P. Tm, A. Pranathi, K. Saiashritha, N. B. Chittaragi, and S. G. Koolagudi, "Tomato Leaf Disease Detection Using Convolutional Neural Networks," 2018 11th International Conference on Contemporary Computing, IC3 2018, Nov. 2018, doi: 10.1109/IC3.2018.8530532.
[31] "LeNet - Wikipedia." https://en.wikipedia.org/wiki/LeNet (accessed Apr. 06, 2023).
[32] G. Sachdeva, P. Singh, and P. Kaur, "Plant leaf disease classification using deep Convolutional neural network with Bayesian learning," Mater Today Proc, vol. 45, pp. 5584–5590, 2021, doi: 10.1016/j.matpr.2021.02.312.
[33] Read, "Bayesian Learning".
[34] A. Smetanin, A. Uzhinskiy, G. Ososkov, P. Goncharov, and A. Nechaevskiy, "Deep learning methods for the plant disease detection platform," AIP Conf Proc, vol. 2377, no. October, 2021, doi: 10.1063/5.0068797.
[35] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Front Plant Sci, vol. 7, no. September, p. 1419, Sep. 2016, doi: 10.3389/FPLS.2016.01419/BIBTEX.
[36] M. Z. Alom et al., "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," 2018, [Online]. Available: http://arxiv.org/abs/1803.01164
[37] Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic study of automated diagnosis of viral plant diseases using convolutional neural networks," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9475, pp. 638–645, 2015, doi: 10.1007/978-3-319-27863-6_59/COVER.
[38] A. V. Panchal, S. C. Patel, K. Bagyalakshmi, P. Kumar, I. R. Khan, and M. Soni, "Image-based Plant Diseases Detection using Deep Learning," Mater Today Proc, Aug. 2021, doi: 10.1016/J.MATPR.2021.07.281.
[39] "Learning rate - Wikipedia." https://en.wikipedia.org/wiki/Learning_rate (accessed Apr. 06, 2023).
[40] S. P. Singh, K. Pritamdas, K. J. Devi, and S. D. Devi, "Custom Convolutional Neural Network for Detection and Classification of Rice Plant Diseases," Procedia Comput Sci, vol. 218, pp. 2026–2040, 2023, doi: 10.1016/j.procs.2023.01.179.
[41] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks," IEEE Access, vol. 7, pp. 59069–59080, 2019, doi: 10.1109/ACCESS.2019.2914929.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:04 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fourth International Conference on Electronics and Sustainable Communication Systems (ICESC-2023)
IEEE Xplore Part Number: CFP23V66-ART; ISBN: 979-8-3503-0009-3

2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC) | 979-8-3503-0009-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICESC57686.2023.10193046

Application of Feedforward Neural Network in Portfolio Optimization and Geometric Brownian Motion in Stock Price Prediction

Anirudh Kumar Aheer
Delhi Technological University
New Delhi, India
anirudh.aheer@gmail.com

Ashish Kumar Pradhan
Delhi Technological University
New Delhi, India
pradhanashish258aug@gmail.com

Dr. R Srivastava
Delhi Technological University
New Delhi, India
rsrivastava@dce.ac.in

Abstract - This research study intends to predict the stock price movement of a set of stocks using the Geometric Brownian Motion and optimize the portfolio built from selected stocks using Feed Forward [1] Neural Network. The onset of the current global pandemic has led to major negative implications for the overall economy and personal financial situations. With many people losing jobs, getting reduced salaries, and exhausting their savings, they have started to invest heavily. The Securities and Exchange Board of India (SEBI) reports that there were 4.9 million new Demat accounts opened in the fiscal year 2020, which was the biggest in at least ten years. Thus, the study of the Stock Exchange is very relevant to the current times. We shall explain the importance of learning about investing in the Stock Market and derive an efficient model to make real-time predictions of a stock price and build an optimized portfolio. In the finance sector, portfolio optimization is a critical responsibility. The goal of portfolio optimization is to increase return on investment while lowering portfolio risk. Due to its capacity to understand intricate correlations between input and output variables, feed-forward neural networks have been successfully used to solve portfolio optimization issues. We explore the research on portfolio optimization using feed-forward neural networks in this paper.

Keywords - Feed Forward Neural Network, Deep Learning, Portfolio Optimization, Geometric Brownian Motion, Stock Market, Machine Learning

I. Introduction

The market trend is the broad direction that the stock is heading in, according to Motilal Oswal Financial Services, a well-known supplier of financial services listed on both the NSE and BSE. Depending on whether the market is bullish or bearish, the trends shift upwards or downwards. Share market trend analysis or equity market trend analysis is the process of looking at current patterns in order to predict future trends. Although there is a lot of data used in this share market trend analysis process, no one can predict the trends with absolute certainty of 100%. Thus, different firms and institutes uses various methods to create future prediction such as chart analysis, technical analysis, and Stock parameters such as Return on Equity (ROE), Price-to-Earning Ratio (PE Ratio), etc to predict the future prices of a particular Stock/Sector.

There are two approaches to reducing risk when making stock investment decisions: gathering the most recent stock market information and diversifying your investments. Investment diversification is the distribution of capital among a number of businesses through the creation of a stock portfolio. A portfolio is a collection of assets that investors own, including both tangible and intangible assets. By building a portfolio of investments, investors can lower their risk of losing money on profits from investments made in the capital market due to stock price volatility. As indicated in the papers, there are several studies on the impact of portfolio optimization on stock prices.

On the capital market, fluctuating stock prices are constantly represented by uncertain stock price movements. The current stock price, which represents the price of the stock in the past, and the market's response to corporate news both affect how the stock price moves, claims the efficient market theory. Stock prices are thought of as random variables, and everyday stock price fluctuations on the capital market are seen as random movements.

One of the major issues in finance is portfolio optimisation. The goal of portfolio optimisation is to distribute the portfolio weights in order to maximise return on investment while lowering portfolio risk. Mean-variance optimisation is one of the traditional portfolio optimisation strategies that is based on the premise that returns have a normal distribution. This presumption might not hold true in reality, and the resulting portfolio might underperform. Due to its capacity to understand intricate correlations between input and output variables, feedforward neural networks have been successfully used to solve portfolio optimisation issues.

The scope of this paper is to conduct a price analysis to predict the movement of future price of selected stocks via application of the Geometric Brownian Motion as the basis of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

the analysis and further use it to build an optimized portfolio for favorable returns.

II. Related Works

[1]"Feed-forward neural networks" by Bebis and Georgiopoulos (1994) introduces the concept of feed-forward neural networks, which are artificial neural networks where the information flows in only one direction, from the input layer to the output layer. The article discusses the structure and functioning of these networks, highlighting their ability to model complex relationships and solve various computational problems. Additionally, it explores the potential applications of feed-forward neural networks in diverse fields such as pattern recognition, signal processing, and control systems. The authors provide insights into the advantages and limitations of this neural network architecture, paving the way for further research and development in the field.

[2]Fama (1995) presents the concept of random walks in stock market prices, which suggests that future stock price movements cannot be reliably predicted based on past price patterns or information. The article explores the implications of this theory for the efficient market hypothesis (EMH) and financial analysis. It discusses how the notion of random walks challenges the traditional methods of technical analysis and fundamental analysis used in stock market forecasting. By examining empirical evidence and market trends, Fama provides valuable insights into the efficiency of stock markets and the role of random price movements, contributing to the understanding of market dynamics and the development of investment strategies.

[3]Telser (1959) proposes a theory of speculation that establishes a relationship between profitability and stability in financial markets. The article argues that speculation serves a vital role in maintaining market stability by aligning prices with fundamental values. Telser introduces the concept of the stabilizing speculation premium, which compensates speculators for their contribution to market stability. By analyzing the interplay between speculative behavior, profitability, and market equilibrium, the author highlights the importance of speculation in reducing price fluctuations and promoting market efficiency. Telser's theory provides a theoretical framework for understanding the dynamics of speculation and its impact on financial markets, offering valuable insights for policymakers and market participants.

[4]Lu, Li, Wang, and Qin (2021) propose a novel method called CNN-BiLSTM-AM for stock price prediction. The approach combines convolutional neural networks (CNNs), bidirectional long short-term memory (BiLSTM) networks, and an attention mechanism (AM) to capture temporal patterns and extract relevant features from historical stock price data. The article describes the architecture and training process of the proposed model, highlighting its ability to capture both short-term and long-term dependencies in the stock market data. The authors present experimental results demonstrating the superior predictive performance of CNN-BiLSTM-AM compared to other traditional and deep learning-based approaches. The findings contribute to the field of stock price prediction and offer insights into the potential of combining different neural network architectures for enhanced forecasting accuracy.

[5]"Louis Bachelier's Theory of Speculation" by Bachelier (2011) provides an in-depth overview of Louis Bachelier's pioneering work on the theory of speculation, which laid the foundation for modern mathematical finance. The article delves into Bachelier's groundbreaking doctoral thesis, "Théorie de la Spéculation," in which he introduced stochastic calculus and developed a mathematical framework to model and analyze stock price movements. Bachelier's work, published in 1900, anticipated future advancements in option pricing, risk management, and quantitative finance. The article discusses the historical significance of Bachelier's theory and its impact on subsequent research and practical applications in the field of finance, cementing his legacy as one of the pioneers of modern financial theory.

[6]Kim and Han (2000) present a genetic algorithm-based approach for feature discretization in artificial neural networks to predict stock price indexes. The article focuses on the preprocessing step of feature discretization, which transforms continuous input variables into discrete values. By employing genetic algorithms, the authors aim to find an optimal set of discretization thresholds that maximize the predictive performance of the neural network model. The article provides a detailed description of the proposed approach, including the encoding of discretization thresholds, fitness evaluation, and the genetic operators used in the optimization process. Experimental results demonstrate the effectiveness of the genetic algorithm-based feature discretization in improving the accuracy of stock price index prediction models, offering valuable insights into feature engineering techniques for financial forecasting tasks.

[7]Reddy and Clinton (2016) investigate the use of geometric Brownian motion (GBM) to simulate stock prices, with a particular focus on Australian companies. The article explores the application of GBM, a stochastic process that assumes the logarithmic returns of a stock follow a normal distribution with constant mean and variance, in modeling stock price dynamics. Reddy and Clinton analyze historical stock price data of various Australian companies and estimate the parameters of GBM to simulate future stock price movements. The findings shed light on the limitations and assumptions of GBM as a modeling tool, providing insights into the accuracy and applicability of this approach for stock price simulation in the Australian market context.

[8]Ma, Han, and Wang (2021) propose a portfolio optimization framework that incorporates return prediction using deep learning and machine learning techniques. The article addresses the challenge of constructing optimal investment portfolios by integrating return prediction models into the portfolio optimization process. The authors introduce a hybrid model that combines deep learning algorithms, such as long short-term memory (LSTM) networks and convolutional neural networks (CNNs), with traditional machine learning

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

methods. The integrated model aims to capture the complex relationships and nonlinearities in financial data to enhance return prediction accuracy. By incorporating the predicted returns into a portfolio optimization algorithm, the authors demonstrate improved risk-adjusted performance compared to traditional approaches. The research contributes to the field of portfolio management by leveraging the power of deep learning and machine learning for enhanced investment decision-making.

[9]Huang, Miao, and Wang (2019) analyze the potential impact of saving China's stock market, examining various factors and policies influencing its stability and growth. The article investigates the challenges faced by the Chinese stock market and explores potential solutions to improve its performance. The authors discuss the role of government interventions, regulatory policies, and investor sentiment in shaping the dynamics of the Chinese stock market. By examining the economic and financial factors influencing stock market stability, the article provides insights into the potential strategies to enhance the resilience and efficiency of the Chinese stock market, which is of significant interest given its size and importance in the global financial landscape.

[10]Jiang (2021) provides an overview of recent progress in applying deep learning techniques to stock market prediction, highlighting advancements and challenges in this area. The article discusses various deep learning architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs), that have been used for stock market forecasting. It explores the utilization of alternative data sources, such as social media sentiment, news articles, and macroeconomic indicators, in combination with deep learning models to improve prediction accuracy. The article also addresses challenges such as data scarcity, model interpretability, and model robustness in the context of deep learning-based stock market prediction. The comprehensive overview of recent developments provides researchers and practitioners with valuable insights into the state-of-the-art techniques and potential future directions in this field. [11]Pang, Zhou, Wang, Lin, and Chang (2020) propose an innovative neural network approach for stock market prediction. The article introduces a hybrid model that combines a convolutional neural network (CNN) and a long short-term memory (LSTM) network to capture both spatial and temporal dependencies in stock market data. The CNN-LSTM model is designed to extract meaningful features from historical stock price sequences and predict future price movements. The authors provide a detailed description of the model architecture, including the convolutional layers, LSTM layers, and the fusion mechanism between the two networks. Experimental results demonstrate the effectiveness of the proposed approach in predicting stock market trends and outperforming traditional time series forecasting models. The research contributes to the field by offering a novel neural network-based approach for stock market prediction that integrates both spatial and temporal information.

[12]Hiransha, Gopalakrishnan, Menon, and Soman (2018) employ deep learning models to predict the National Stock Exchange (NSE) stock market in India. The article explores the application of deep learning techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), in modeling and predicting stock market movements. The authors preprocess the financial data and train the deep learning models using historical stock price and trading volume data. The article provides insights into the performance and accuracy of the deep learning models in forecasting the NSE stock market, highlighting the potential of these models for predicting financial market trends. The findings contribute to the understanding of deep learning applications in stock market prediction and provide practical implications for investors and financial analysts in the Indian stock market context.

[13]Mukherjee, Sadhukhan, Sarkar, Roy, and De (2023) investigate the use of various deep learning algorithms for stock market prediction, comparing their performance and analyzing the underlying mechanisms. The article explores popular deep learning architectures such as convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and hybrid models in the context of stock market forecasting. The authors conduct experiments using historical stock price data and evaluate the prediction accuracy of different deep learning algorithms. They analyze the strengths and limitations of each approach, considering factors such as model complexity, interpretability, and computational efficiency. The research provides valuable insights into the selection and optimization of deep learning models for stock market prediction, offering guidance for researchers and practitioners in the field.

[14]Mintarya, Halim, Angie, Achmad, and Kurniawan (2023) conduct a systematic literature review on machine learning approaches in stock market prediction. The article synthesizes and analyzes existing research on the application of machine learning techniques for stock market forecasting. The authors identify common machine learning algorithms used in this domain, such as support vector machines (SVM), random forests, and neural networks. They examine the datasets, features, and evaluation metrics employed in different studies. The review provides an overview of the strengths and limitations of machine learning approaches in stock market prediction, highlights trends, and identifies potential areas for future research. The comprehensive review contributes to the understanding of the current state of machine learning in stock market prediction and serves as a valuable resource for researchers and practitioners interested in this field.

[15]Brătian, Acu, Mihaiu, and Șerban (2022) analyze the behavior of stock indexes under geometric Brownian motion (GBM) and investigate the relationship between financial market uncertainty and non-crisis as well as financial crisis scenarios. The article explores the application of GBM, a widely used stochastic process in finance, to model the dynamics of stock indexes. The authors simulate stock price movements using GBM and analyze the impact of different levels of market uncertainty on the behavior of stock indexes

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

III. Objective

Our aim is to conduct a price analysis for multiple stocks listed on the NSE, which in our case are City Union Bank, Spicejet, Astrazeneca Pharma, Wockhardt, jubilant pharmova, Tata Motors, ITC, HDFC , Ultratech Cements and Sun Pharmaceuticals using the Geometric Brownian Motion. We will analyze the Stock price for a time frame using the programming language Python. We will first pre-process the data (clean the data and remove all values unnecessary for the scope of this analysis) and then implement the Geometric Brownian Motion Model to conduct a comparative analysis of the stock prices. Then we will implement Feed Forward [1] Neural Network implemented with a single hidden layer using the programming language Python to create an optimized portfolio out of stocks with better returns.

IV. Brownian Motion

More than just the flow of particles out of a liquid, Brownian motion has been applied mathematically as well.
A standard Wiener process also referred to as standard Brownian is a stochastic process.

$$W_t(t >= 0); \tag{1}$$

which represents, a family of random variables W_t , defined on a similar probability space (Ω,F,P)) and indexed by non-negative real values t, having properties as following:

- W_o = 0
- The function t $\rightarrow W_t$ is continuous in t, with probability 1,
- The increments of the process $W_t, t \geq 0$ are stationary and independent.
- The distribution of the increment $W_{(t+s)}$ - W_s is normal(0,t).

Independent increments refer to the property that for each choice of real numbers which are non-negative,

$$0 \leq x1 < y1 \leq x2 < y2 \leq \ldots \leq xn < yn < \infty, \tag{2}$$

the random variables represented by the increments

$$W_{y1}–W_{x1}, W_{y2}–W_{x2}, ..., W_{yn}–W_{xn} \tag{3}$$

are independent jointly;
By stationary increments, it is meant that for any

$$0 < x, y < \infty \tag{4}$$

the increment's distribution $W_{y+x}–W_x$ should follow the same distribution as $W_t–W_0 = W_t$.

V. Geometric Brownian Motion

'Brownian Motion' can have negative values, making it dubious to use it directly to simulate stock prices among other things.Consequently, we use the 'Geometric Brownian Motion', a non-negative variant of Brownian Motion:
S(t), which can be defined as

$$S(t) = S0 * e^{X(t)} \tag{5}$$

Where,

$$X(t) = \sigma * B(t) + \mu * t \tag{6}$$

is Brownian Motion which has drift as well as volatility.

$$S(0) = S_o > 0 \tag{7}$$

represents the initial value.
To determine the Brownian Motion, logarithms are used;

$$X(t) = ln[S(t)/S_o] = ln[S(t)) - ln(S_o]. \tag{8}$$

$$\ln(S(t)) = \ln(S_o) + X(t) \tag{9}$$

is normal with mean: μ*t + ln(S_0), and variance: $\sigma^2 * t$; hence, S(t) has a log-normal distribution for every t.
OR

$$S_{pred} = S_o * e^{(drift+diffusion)} \tag{10}$$

Where,
So: The closing price of the beginning time point
Drift: the longer-term trend in stock prices is shown.
Diffusion/Volatility: reflects shorter-term fluctuations

Two components make up GBM: a definite component and an uncertain component. A specific/definite component reflects the return on the stock over a limited period of time and is also referred to as the stock's drift. The uncertain component is a stochastic process that depicts the randomness of the stock price with a component of random volatility and a component of stock market volatility.

A. Drift

- The longer-term trend in stock values can be seen in drift. As you can see in the expression above, drift is a constant value that may be applied to a certain time point's stock price (t-1) if we assume there is no random shock.
- Stock prices will gradually increase if the drift is favorable or decrease if the drift is unfavorable as long as there is no sudden shock. We can see that because drift is applied using the exponential function, stock prices can never reach zero.

B. Volatility/Diffusion

- Diffusion reflects transient changes. We multiply the sigma by the random value z(k) to determine the diffusion component.
- Several stock price prediction scenarios can be created thanks to the diffusion component. The component of drift is constant.It limits the number of scenarios we can create, but the diffusion component, which uses the Wiener process to create independent, stationary, and evenly spaced random shocks, allows us to create as many scenarios as we want.
- Therefore, we must add the effects of both components when making a prediction for the time point (k) while

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

taking into account the stock price at the time point (k-1). We can make a prediction about what will happen at time point (k) using the combined total effect.
- We could always predict the stock price at the following time point if there was only the drift component, but the diffusion component disturbs up this smoothness by injecting random shock. As a result, neither a steady upward trend nor a downward trend in stock prices can be seen. .

The GBM model assumes that stock returns are normally distributed and that stock prices follow a log-normal distribution.
The model is based on three parameters:

- the price of stock at the current time (S_o)
- the expected rate of return (μ), and
- the volatility (σ) of the stock.

The following equation is what the GBM model predicts the stock price will follow:

$$dS = \mu S dt + \sigma S dW \tag{11}$$

where dW is a Wiener process or Brownian motion, dS represents the shift in stock's price, dt represents the change in time, μ symbolizes the expected rate of return and σ is used to represent volatility.

VI. Feed Forward Neural Network

A variant of Artificial Neural Network (ANN) called as Feed-Forward Neural Network (FFNN) only permits data to move from the initial input layer towards the output layer.The output layer creates the finished product after getting the input data. In order to retrieve pertinent information between the input and output layers, the hidden layers execute computations on the input data.

Every neuron in a layer of a feed-forward neural network is connected to every other layer below it, but there exist no connections between layers that are higher or lower or within the same layer. While each neuron comprising the input layer represents a feature of the input data, each neuron comprising of the output layer generates a prediction or classification based on the input data.

Each neuron in the feed-forward neural network is given weights and biases, which are changed during training to reduce the difference in the discrepancy between the anticipated and actual production. The biases and weights govern the threshold for neuron activity and the strength of the connection between neurons, respectively. Classification, regression, and pattern recognition are just a few of the tasks that feed-forward neural networks are capable of. Gradient descent-based algorithms can efficiently train them, and they are easy to implement. However, they might not be appropriate for difficult tasks that call for a lot of data or have a lot of fluctuation in the input data.

VII. Feed Forward Neural Network with Single Hidden Layer

A feed-forward neural network with one hidden layer is an Artificial Neural Network (ANN) comprising an input layer, a hidden layer, and an output layer. Data is first received by the input layer, then processed by the hidden layer, and finally produced by the output layer.

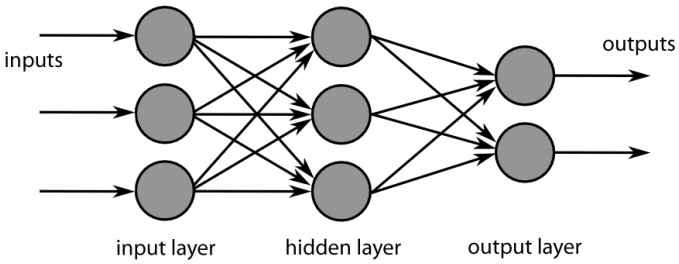


Fig.1 FFNN with single hidden layer

The circles in the above diagram stand in for individual neurons, while the lines joining them stand in for their connections. Three neurons form the input layer, the middle section of four neurons makes up the hidden layer, and the output layer is comprised of the two neurons.
The input layer receives the data as input, and the output is given to the hidden layer. The hidden layer in a feedforward neural network is significant as it extracts important features, enables non-linear mapping, processes information, facilitates generalization, and determines the network's capacity to learn and represent complex relationships. It acts as a feature extractor, transforming input data into abstract representations. The non-linear transformations in the hidden layer allow the network to capture complex patterns. It serves as a computational unit that processes information and helps the network generalize from training data to unseen examples. The number of neurons and layers in the hidden layer affects the network's capacity and expressiveness. The output is then created by the hidden layer by performing a linear combination of the inputs along with the application of an activation function, such as the sigmoid function. The output layer then receives this output and creates the network's final output.

Instead of a step function a single-layer neural network can compute a continuous output. The 'logistic function' is a well-liked choice.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{12}$$

By making this decision, the single-layer network is changed into a model that is exactly like the logistic regression model, a popular statistical modeling technique. The logistic function is a member of a group of mathematical operations known as sigmoid functions because of the way their S-shaped graphs resemble the Greek letter Sigma in lowercase. Owing to the fact that the derivative of this function is continuous, it can be applied in backpropagation. An additional reason to choose

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

this function was that it is straightforward to calculate its derivative

$$f'(x) = f(x)(1 - f(x)) \tag{13}$$

VIII. Data

To ensure the accuracy and relevance of our research paper on the stock market, we obtained access to reliable stock prices and related information from the National Stock Exchange (NSE) website. We used the yfinance package to analyze data on the stocks of Sun Pharmaceutical Industries, Spice Jet, Jubilant Pharmova, ITC Ltd, HDFC, City Union Bank, AstraZeneca Pharma India Limited, Wockhardt Limited, Tata Motors Limited, and UltraTech Cement Limited. These companies represent a diverse range of industries and sectors. For instance, Sun Pharmaceutical Industries is a prominent Indian pharmaceutical company, while Spice Jet is a well-known Indian airline company. AstraZeneca Pharma India Limited is a multinational pharmaceutical firm with a focus on the discovery, production, and research of novel medications. A well-known public sector bank in India, City Union Bank provides a variety of financial goods and services. An Indian pharmaceutical company called Wockhardt Limited produces and sells a wide variety of medical supplies. Commercial and passenger cars are produced by the Indian multinational automobile manufacturer Tata Motors Limited. UltraTech Cement Limited is an Indian cement company that manufactures and markets various types of cement and building materials. By analyzing stock prices and other pertinent data, we aim to gain valuable insights into market trends and investor behavior. The collection of data from the NSE website via yfinance constitutes a critical step towards conducting a comprehensive and insightful research paper on the stock market.

IX. Methodology

We have followed a step-wise approach which begins with understanding our data set and what it is composed of. We first import consolidated CSV files separately containing the stock prices of the mentioned stocks for the period 17th March 2020 to 17th March 2023 from a database of NSE India using finance in the programming language Python. We further removed the columns that were unnecessary for our analysis. As the next step, we create a model to implement the Geometric Brownian motion and create visualizations to compare the results. Having done that, we choose a set of stocks with better returns as predicted by GBM for the creation of an optimized portfolio. We implement Feed Forward Neural Network with a single hidden layer using the programming language Python on the Google Colab platform. We import Stock price data of the selected stocks for the last 10 years to train the model. Further, to compare the returns, we implement visualizations.

X. Application of GBM Model

The Model for implementing the Geometric Brownian Motion has been created using Python as the Programming Language and executed on the platform Google Colab. The parameters used for the application of model are:

1) T - Total Time period for prediction
2) dt - Incremental time period between two consecutive closing prices,in our case it is 1 day
3) N - Number of time points in the prediction time horizon, N = T/dt
4) t - The time progression in our model is displayed in an array. We can measure time by counting the number of time points that have passed, just like on a time ticker.
5) μ - The average return of the stock prices over the historical period we are predicting.
6) σ - The returns from previous stock prices' standard deviation.
7) b - We introduce randomization to our model using an array. Contains a random number drawn from the normal distribution of the standard distribution for each relevant predicted time point. The random shocks will be added by these random numbers.
8) W - From the time point at the starting (So) to another time point, values of the stock will change according to the Brownian path.
9) S_o - The closing price of the beginning time point, in our case it is the Closing price of day 0

For analysis, we take the array of returns;

$$R_t = (S_t - S_{(t-1)})/S_{t-1} \tag{14}$$

R_t: Stock return at time t
S_t : the value of a stock at time t
$S_{(t-1)}$: Cost of a stock on the previous day (t–1)

Now, for calculation of μ and σ:
μ = mean of return of the stock prices
σ= standard deviation of return of stock prices
We calculated the expected rate of return (μ) and volatility(σ) using the 80% of data from 17/03/2020 to 17/03/2023 split into 80:20 ratio, and use it to forecast the stock movement for the remaining 20% of data.

$$drift = \mu - 1/2\sigma^2 t \tag{15}$$

We compare the stocks using mean return during the Forecast period.

XI. Outcomes of GBM Modelling

The simulation results of the Geometric Brownian Motion model for the Indian stocks suggest that the model can provide reasonable predictions of the future prices of these stocks in the Indian market. The model accurately captured the general trend of the stock prices over time and showed that the prices tend to fluctuate around the mean with a certain level of volatility, which is a common phenomenon in the Indian stock market.

These charts compare the Forecasted value vs the Actual value for these stocks: Here in the figure, the Blue Line represents the Forecasted values of the GBM model over time

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

and the orange line shows the actual fluctuation in the stock prices.

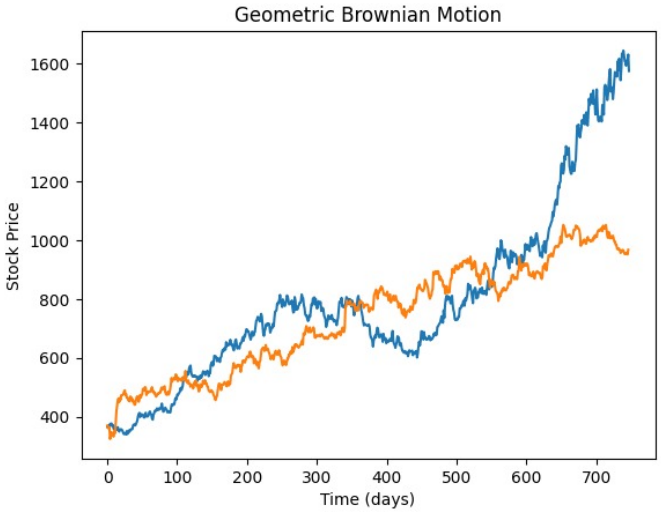


Fig.2 Sun Pharma Stock price prediction

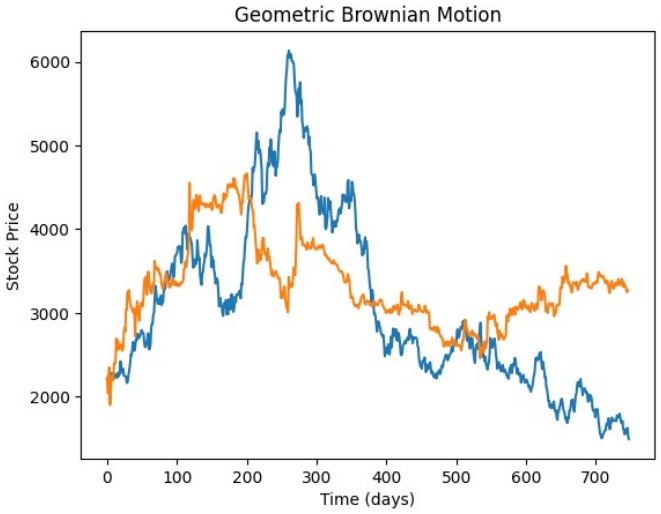


Fig.3 Astrazeneca Stock price prediction

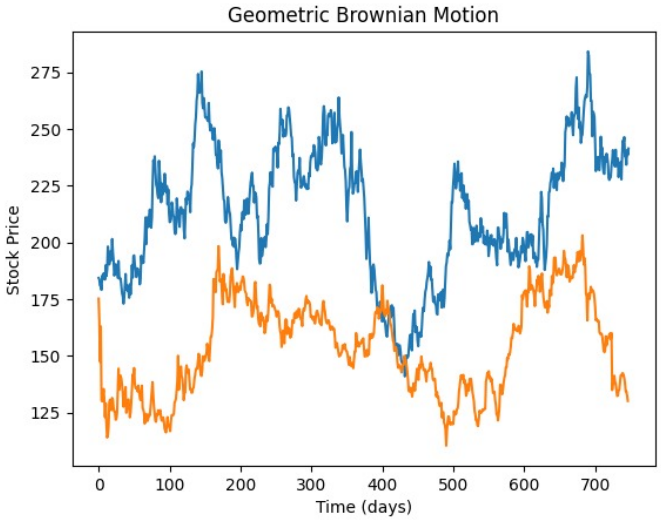


Fig.4 City Union Bank Stock price prediction

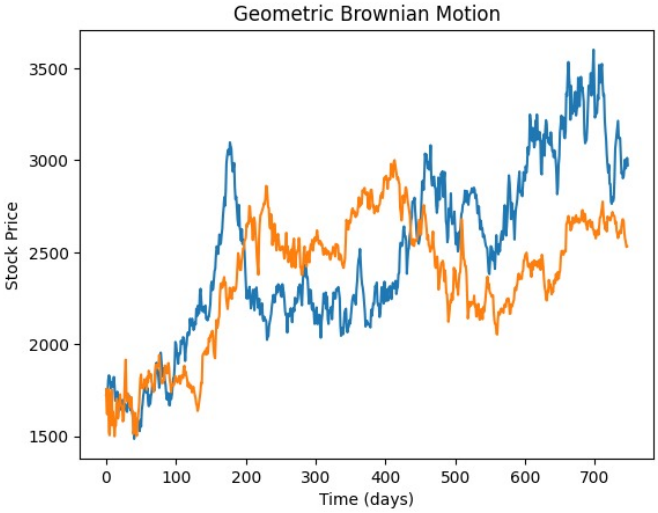


Fig.5 HDFC Bank Stock price prediction

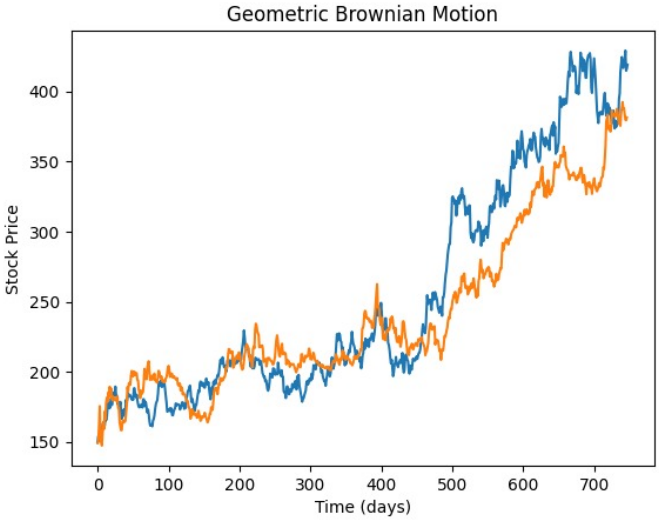


Fig.6 ITC Stock price prediction

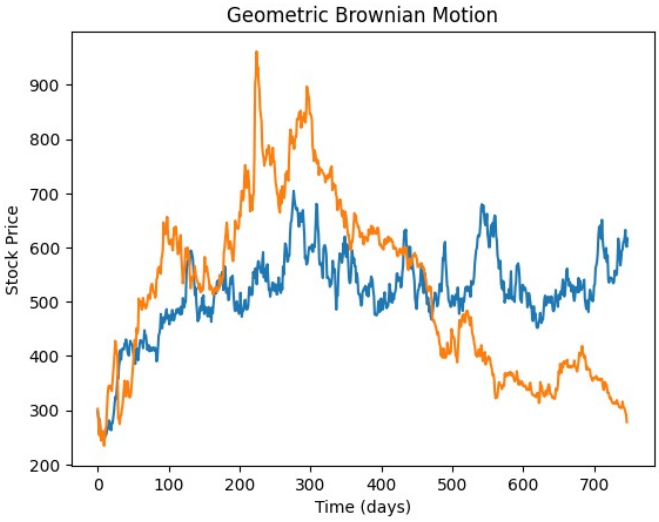


Fig.7 Jubilant Pharmova Stock price prediction

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

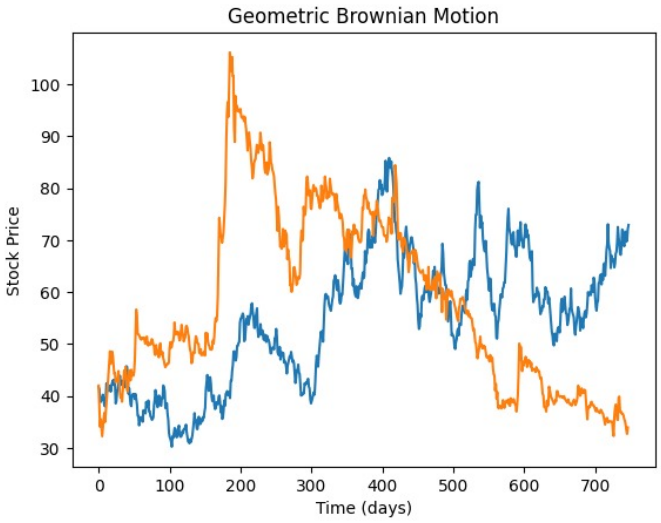


Fig.8 Spicejet Stock price prediction

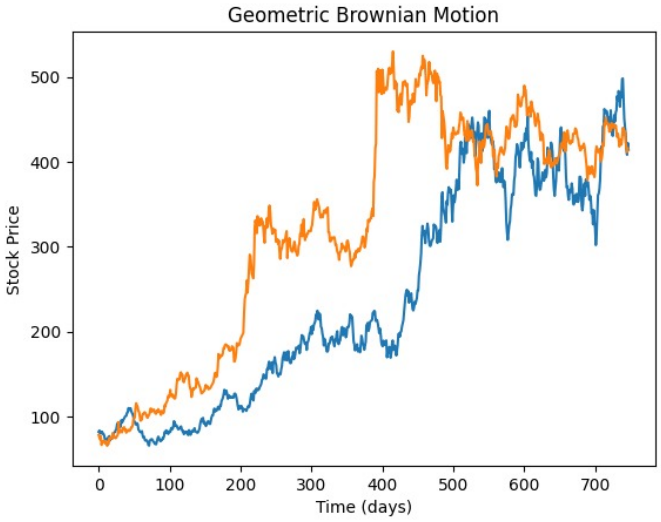


Fig.9 Tata Motors Stock price prediction

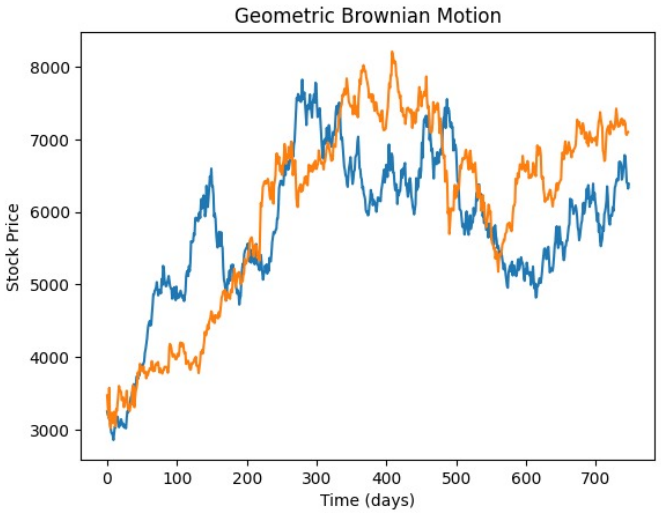


Fig.10 Ultratech Cements Stock price prediction

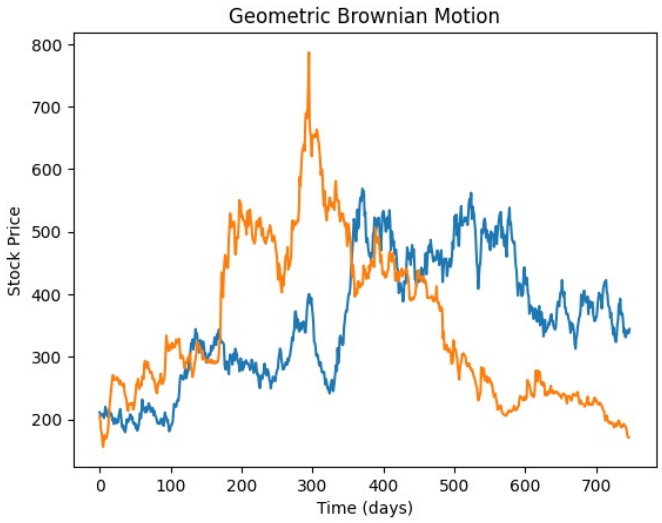


Fig.11 Wockhardt Ltd. Stock price prediction

We now compare the expected return on stocks through predicted values over a period of six months:

S No.	Stock	Return
1	SUNPHARMA	0.003378
2	City Union Bank	-0.001497
3	SPICEJET	0.002355
4	Astrazeneca	0.000542
5	Wockhardt	0.002105
6	Jubilant Pharmova	0.004663
7	TATA Motors	0.000608
8	ITC	0.001397
9	HDFC	0.000794
10	Ultratech Cements	0.001279

XII. Application of Feed Forward Neural Network

We start with selecting our data and then we calculate the daily returns. We define the inputs for the neural networks. We then split out the entire data into a ratio of 80:20 Then we create our training and testing data. After defining the structure of our model we start training the model. We use mean square error as a validation method to notably find out the loss and the error.

XIII. Outcomes of Portfolio Optimization using Feed Forward Neural Network

We trained the model using a large data set of historical stock market data and observed that the training loss consistently decreased during the initial training period. However, after the 20th epoch, the training loss reduction rate slowed down significantly, indicating that the model had reached a certain level of optimization.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

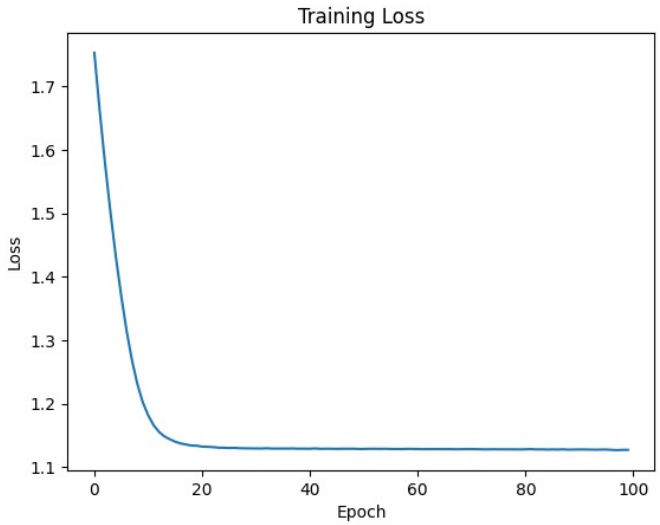


Fig.12 Training Loss

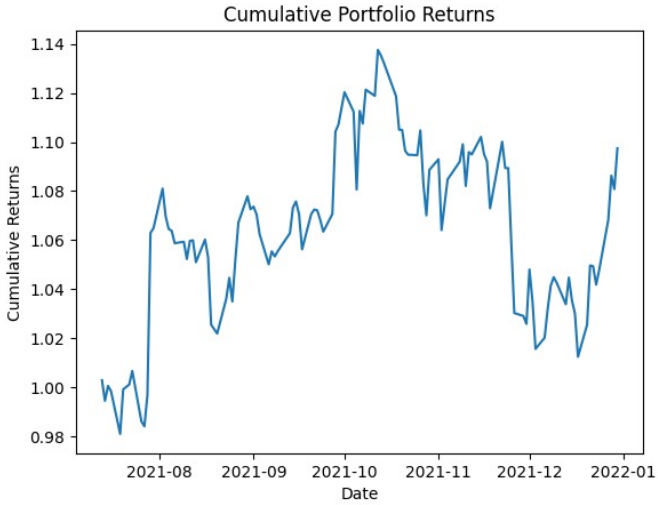
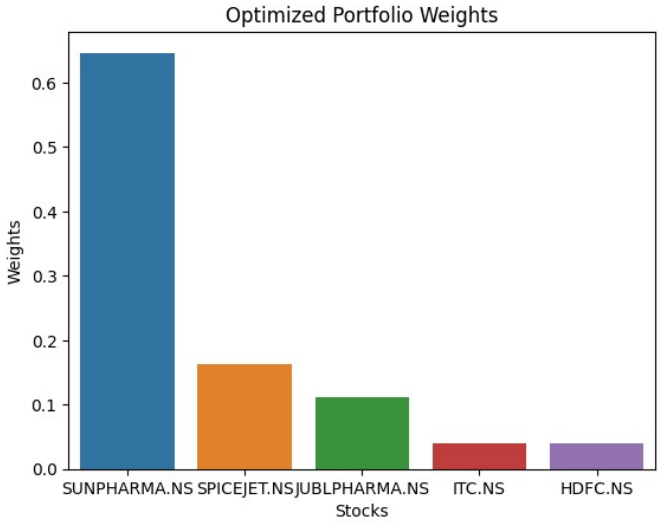


Fig.13 Cumulative Portfolio Returns

Optimized portfolio Weights as calculated using Feed Forward Neural Network with a single hidden Layer after training the model is:



Optimized Portfolio Weights

S No.	Stock	Optimized Weight
1	SUNPHARMA	0.6283615
2	SPICEJET	0.17370817
3	JUBILPHARMA	0.11454658
4	ITC	0.04296229
5	HDFC	0.04042152

XIV. Conclusion and Future Work

The model performs well only when the following conditions pertaining to the stock prices are satisfied:

- The businesses are still operating, and their stock prices are stable over time and in terms of value.
- Because stocks show the pattern of a Markov process, it is only possible to predict future prices using the current stock price.
- A stock's proportional return has a log-normal distribution.
- A stock's continually compounding return is typically distributed.
- The presence of such mathematical constraints limits the application and generalization of the GBM model in the actual Stock investment and/or trade analysis.

Through our analysis we can conclude if the above conditions are met, then the GBM model for stock simulation can be used to predict the future prices of the selected Stock.
However, the model also had some limitations in accurately predicting the exact prices of the stocks, particularly during periods of high volatility. Overall, the results of this study suggest that the GBM model can be a useful tool for predicting the future prices of Indian stocks in the Indian market, but it should be used in conjunction with other forecasting techniques for more accurate predictions.

In addition to the findings discussed above, our study also employed a portfolio optimization technique using a feed-forward neural network on the five stocks selected out of the ten. The results showed that the neural network approach

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

outperformed the GBM model in predicting future stock prices and generating maximum returns on investment. Furthermore, the portfolio optimization technique allowed us to identify the best possible combination of stocks to invest in for maximizing returns. Therefore, our study suggests that a combination of the GBM model and portfolio optimization techniques, such as neural networks, can be a powerful tool for Stock Trading or Swing trades.

REFERENCES

[1] Bebis, G., Georgiopoulos, M. (1994). Feed-forward neural networks. Ieee Potentials, 13(4), 27-31.
[2] Fama, E. F. (1995). Random walks in stock market prices. Financial analysts journal, 51(1), 75-80.
[3] Telser, L. G. (1959). A theory of speculation relating profitability and stability. The Review of Economics and Statistics, 295-301.
[4] Lu, W., Li, J., Wang, J., Qin, L. (2021). A CNN-BiLSTM-AM method for stock price prediction. Neural Computing and Applications, 33, 4741-4753.
[5] Bachelier, L. (2011). Louis Bachelier's Theory of Speculation. In Louis Bachelier's Theory of Speculation. Princeton University Press.
[6] Kim, K. J., Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of a stock price index. Expert Systems with Applications, 19(2), 125-132.
[7] Reddy, K., Clinton, V. (2016). Simulating stock prices using geometric Brownian motion: Evidence from Australian companies. Australasian Accounting, Business and Finance Journal, 10(3), 23-47.
[8] Ma, Y., Han, R., Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications, 165, 113973.
[9] Huang, Y., Miao, J., Wang, P. (2019). Saving China's stock market?. IMF Economic Review, 67, 349-394.
[10] Jiang, W. (2021). Applications of deep learning in stock market prediction: recent progress. Expert Systems with Applications, 184, 115537.
[11] Pang, X., Zhou, Y., Wang, P., Lin, W., Chang, V. (2020). An innovative neural network approach for stock market prediction. The Journal of Supercomputing, 76, 2098-2118.
[12] Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., Soman, K. P. (2018). NSE stock market prediction using deep-learning models. Procedia computer science, 132, 1351-1362.
[13] Mukherjee, S., Sadhukhan, B., Sarkar, N., Roy, D., De, S. (2023). Stock market prediction using deep learning algorithms. CAAI Transactions on Intelligence Technology, 8(1), 82-94.
[14] Mintarya, L. N., Halim, J. N., Angie, C., Achmad, S., Kurniawan, A. (2023). Machine learning approaches in stock market prediction: A systematic literature review. Procedia Computer Science, 216, 96-102.
[15] Brătian, V., Acu, A. M., Mihaiu, D. M., Șerban, R. A. (2022). Geometric Brownian Motion (GBM) of stock indexes and financial market uncertainty in the context of non-crisis and financial crisis scenarios. Mathematics, 10(3), 309.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:36:29 UTC from IEEE Xplore. Restrictions apply.

Water Air Soil Pollut (2023) 234:571
https://doi.org/10.1007/s11270-023-06569-1

Assessment of Spatiotemporal Variations in Water Quality of the Urban River Reach, Yamuna, Delhi

Nibedita Verma · Geeta Singh · Naved Ahsan

Received: 22 November 2022 / Accepted: 31 July 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract Yamuna River Delhi stretch is a leading source of water supply to Delhi. The river stretch is a receptacle of urban liquid waste of human interference from domestic and industrial fields and conduces to the most contaminant reach of the country. The present study used multivariate techniques to represent the spatiotemporal water quality variations and interpreted a large hourly complex dataset (March 2021–February 2022) obtained from two real-time monitoring stations situated upstream and downstream of this reach. Eleven water quality parameters were assessed, and Box Whisker plots were drawn monthly basis. The increased concentrations of conductivity, BOD, COD, TOC, NH4, and low DO downstream indicated the influence of outfalling drains and diffused sources contributing pollutants into the river stretch. A higher BOD, COD, and TOC concentration was observed downstream in monsoon attributed to the organic substance in surface runoff. FA and PCA were implemented in the standardized data set to reveal the correlation between the water parameter. For upstream, turbidity, TOC, COD, and TSS have strongly positively loaded for factor 1 for all the seasons except monsoon. For downstream. TOC and COD contributed strongly positive load except in winter. The study reveals that the urban river flows with agricultural and surface runoff: industrial and domestic wastewater with organic and inorganic substances. This present study is a novel approach to assessing spatial and temporal variation of an urban reach for the current geomorphological conditions with real-time data.

Keywords Water quality variations · Urban river reach · Box Whisker plot · Principal component analysis · Factor analysis

1 Introduction

Rivers are the most sensitive natural resources as they carry domestic, industrial, and agricultural runoff from the large watershed (Zhang et al., 2009). The river conditions rely on natural and anthropogenic activities such as precipitation, erosion, municipal waste, land use pattern, urbanization, industrialization, agricultural flow, and mining (Yilma et al., 2019). The combined effect of these activities deteriorates the surface water bodies, which need (Wang et al., 2013) appropriate management options. Vegetation, topography, geological characteristics of the

N. Verma · G. Singh (✉)
Department of Environmental Engineering, Delhi Technological University, New Delhi, Delhi 110042, India
e-mail: geeta.singh@dce.ac.in

N. Verma
e-mail: nibedita_2k19phd@dtu.ac.in

N. Ahsan
Civil Engineering Department, Jamia Millia Islamia University, Jamia Nagar, New Delhi 110025, India
e-mail: nahsan@jmi.ac.in

Published online: 19 August 2023

catchment area, seasonal variation of the precipitation levels, surface runoff, ground surface interaction, interception, and abstraction promote the water quality parameter of a water body (Avtar et al., 2013). The spatiotemporal assessment of river water quality requires regular and continuous monitoring programs to estimate and maintain (Singh et al., 2004). This monitoring leads to the creation of vast and complex data matrices composed of many physicochemical constituents, which are challenging to interpret and conclude (Bonansea et al., 2018; Singh et al., 2004). A reliable river quality assessment framework is required to minimize the impacts of pollutants and protect a river's ecological health and sustainable expansion (Arora & Keshari, 2021). Nowadays, principal component analysis (PCA), and factor analysis (FA), are widely accustomed to assess river characteristics by simplifying complex data and removing spare information, and keeping valuable information (Bonansea et al., 2018).

Multivariate statistical techniques create a matrix from acquired data to assess relation, likeness, and unlikeness (Singh et al., 2004) among the constituents and their origin (Achieng et al., 2017). The PCA and CA (Cluster Analysis) help to distinguish potential pollutant sources affecting water resources and provide a unique understanding of the relationship between different pollutants (Singh et al., 2020; Wang et al., 2013). These methods are essential to verifying spatiotemporal variation generated by various natural and manmade factors (Shrestha & Kazama, 2007; Vega et al., 1998). PCA/FA is used to assess river quality, identify pollutant sources, and as a tool for water resource management (Atakoglu & Yalcin, 2022; Dehghanzadeh et al., 2015). PCA reduces the dimensionality of factors and keeps all related information of corresponding constituents (Singh et al., 2004). Achieng et al. (2017) deployed multivariate statistics to interpret the seasonal variation of the Sosiani River, Kenya, and evaluated the effects of anthropogenic activities. FA explained the specific contamination sources in the Karstic region (Calijuri et al., 2012). The hydrogeochemical influences and anthropogenic activities on groundwater were assessed for the Ken-Betwa River (Avtar et al., 2013). Setia et al. (2021) investigated the Sutlej River water quality and concluded the presence of ions in transboundary regions with the help of multivariate techniques PCA/FA and CA successfully evaluated the spatial variations in river quality and identified hidden contributors of contaminants in the Songhua River (Wang et al., 2013). Bu et al. (2010) applied multivariate statistics to appraise the Jinshui River quality in China and classified the sites with 22 variables into three clusters. They extricated five factors revealing 90.01% of the total variance by identifying the sources for watershed management. A case study of the Nakdong River, Korea, with multivariate tools, revealed that manmade activities controlled the river quality (Venkatramanan et al., 2014). Singh et al. (2020) examined the relationship between physicochemical variables with pollution index and PCA to assess pollutant sources suggesting management procedures for Kali River, India. PCA/FA, DA (discriminant analysis), and multiple regression analysis are used to reduce complex data sets and identify the factors and sources of pollutants for the high-low flow conditions of the Melen River, Turkey (Koklu et al., 2010). Githaiga et al. (2021) applied water pollution evaluation indices to assess heavy metals' water quality, concentration, and health risks. They also used PCA and CA to deduce the sources of these metals. PCA and CA were used no and locations of monitoring stations for the Lis River (Vieira et al., 2012). Bonansea et al. (2018) estimated the reservoir water quality for eleven sampling sites of two reservoirs in Argentina. In contrast, cluster analysis grouped samples into three clusters, and PCA/FA shows the existence of five. PCA explained the groundwater quality of Guwahati City of Assam and determined that rock weathering is the primary factor for the pollution of groundwater (Hinge et al., 2022). Artificial activities were established by using PCA and Pearson's correlation for the Marreco River (Priya & Arulraj, 2011).

The present study applied PCA, FA, and Box Whisker plots for analyzing the real-time dataset of the urban reach of the Yamuna River, Delhi, India. This stretch has been turned into a sewage drain due to tremendous anthropogenic pressure aggravating the pollution load (Upadhyay et al., 2011). The stretch between Wazirabad to Okhla gets massive pollutant input from eighteen major drains comprising domestic sewage and industrial effluent from different clusters of Delhi (CPCB, 2012). Khullar and Singh (2022) assessed water quality with a deep learning method for the Palla, which is upstream of the Delhi reach, Arora and Keshari (2021) used a pattern recognition method to assess the river stretch

with the five years data for the year 2006–2010 for seven parameters. Sharma and Kansal (2011) evaluated the water quality index for the monthly data of 1999–2009 for BOD (Biochemical Oxygen demand) and DO (Dissolved Oxygen). The objectives of the present study were to evaluate the spatiotemporal variations of eleven physiochemical water quality parameters from the large one-year hourly datasets (March 2021–February 2022) for the upstream and downstream of the Yamuna River Delhi stretch. The study revealed the hidden factors describing the large data sets and impacts of probable water pollutant sources. The present study assessed the variation of water quality parameters for current geomorphological and meteorological conditions to administrate a sustainable water quality management commencement. This study is a novel approach to assessing spatial and temporal variation upstream and downstream of an urban reach for real-time data sets.

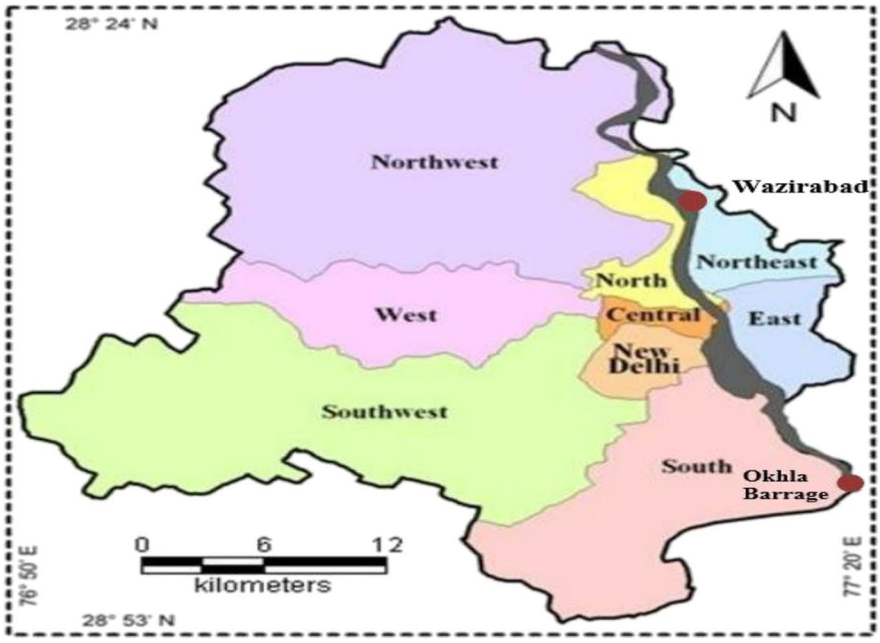
2 Materials and Methodology

2.1 Study Area

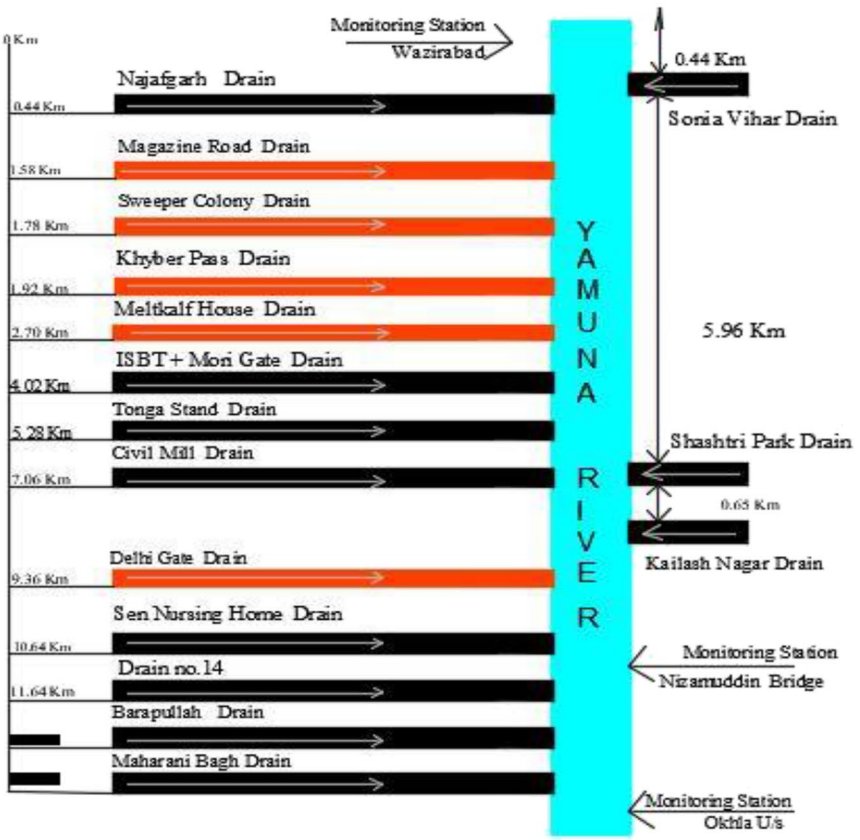
Originating from Yamunotri, located at the coordinates 31° 01′ N and 78° 46′ E (CPCB, 2006), Yamuna River, the lengthiest tributary of the Ganges, travels 1376 km and joins the Ganga River at PrayagRaj. Before entering Delhi, it covers 348 km from its origin. The Yamuna River enters Delhi 1.5 km upstream of Palla village, travels 48 km to reach Okhla, and leaves at Jaitpur downstream of Okhla village(Upadhyay et al., 2011). The urban reach of the Yamuna, Delhi, is located between Wazirabad (28°49′24.39″N and 28°31′50.99″ N) and Okhla (77°13′39.92″ E and 77°20′36.8″ E), flowing 22 km long (CPCB, 2006). Figure 1(a) shows the Delhi reach of the Yamuna River from Wazirabad (ST1) to Okhla (ST2), which receives untreated and partially treated wastewater from 16 main drains shown in Fig. 1(b) carrying domestic sewage and industrial effluents. One of the largest drains, Shadra Drain, outfalls downstream of the Okhla barrage (CPCB, 2012), is excluded from this study. The reach from Palla to Wazirabad, for the considerable flow, the water contains adequate DO and low BOD (Arora & Keshari, 2021). After Wazirabad, scarce fresh water is available due to the water abstraction at Wazirabad for the water supply of Delhi. The outfalling drains generate 3393 MLD of wastewater with 229 TPD BOD loads per day (CPCB, 2012). Low flow and outfalling wastewater with high organic and nutrient loads cause the river conditions to degrade (Mandal et al., 2010). The wastewater from Hindon also disposes of upstream of Okhla (Arora & Keshari, 2021). Delhi is in a subtropical zone with the highest temperature of around 45 ºC and an average temperature of 32ºC. During winter lowest temperature is 2ºC, and the average is 12ºC. Monsoon approaches late June and ends in mid-September. For Delhi, rainfall is highest (234 mm) in August and contributes 32% of annual rainfall (715 mm) (Rana et al., 2012). Hence, dilution is higher during the monsoon period. The data were collected from two Real-Time Water Quality Monitoring Stations (RTWQMS) of the Delhi Pollution Control Committee (DPCC) upstream of the Wazirabad (ST1) barrage and upstream of the Okhla (ST2) barrage.

2.2 Data Used and Methodology

The present study collected data from the Real-Time Water Quality Monitoring Stations (RTWQMS) of the Delhi Pollution Control Committee (DPCC) at Wazirabad and Okhla. These RTQWMSs use S::CAN water monitoring technologies. UV Spectro::lyser is used to determine TSS (total suspended solids), turbidity, BOD (biochemical oxygen demand), COD (chemical oxygen demand), TOC (total organic carbon), and NO3-N (nitrate-nitrogen). Ammino::lyser Eco is used to measure Ammonia. pH::lyser is used to monitor pH and Temperature. The data has been collected from ST1 (Wazirabad) (6085 hourly observations) and ST2 (Okhla) (25,540 no observations at 15-min intervals). The observation period was March 2021 to February 2022 with current geomorphological conditions. The parameters analyzed were DO, BOD, COD, pH, water temperature, conductivity, turbidity, ammonium-nitrogen (NH4), nitrate nitrogen (NH3), total suspended solids (TSS), and total organic carbon (TOC). Hourly data had been used for multivariate analysis and extracted valuable information. Collected data were subdivided monthly to evaluate the average monthly water quality variations by Box Whisker plots. For PCA/FA, complex datasets were divided into four subdivisions – summer, monsoon, post-monsoon, and winter. Box Whisker plots



(a)



(b)

Fig. 1 **(a** & **b)** Yamuna River at Delhi with a line diagram showing monitoring stations (ST1 and ST2) and a line diagram showing respective drain positions (Source: DPCC)

were used to interpret the temporal and spatial variation in 11 parameters. Significant parameters have introduced a classification for further assessment. The data sets were subdivided monthly basis, and Box Whisker plots were drawn for each parameter and both stations to interpret the spatiotemporal variations of parameters, which are shown in Figs. 2 and 3.

The PCA is a data-minimizing method (Singh et al., 2020). It gives details of the most crucial constituents explaining the variability of data sets (Singh et al., 2004). PCA reshapes the indigenous variables into a lower set (Vega et al., 1998) and can be expressed as (Bu et al., 2010)

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \cdots + a_{im}x_{mj} \quad (1)$$

where z and a are component score and loading, respectively, x = observed variables quantity. i = constituent no, j = sample no, m = total no of variables. FA was computed using varimax rotation generating factors. PCA is included in FA, and FA is dominating tool to lower the dimensionality of the data matrix with a high no of related variables (Singh et al., 2004). The PCs are extracted from the cross-product data matrix and help to explain the disintegration of multiple parameters by obtaining eigenvalues and eigenvectors. Varifactors or factors obtained after rotating Varimax rotation distributes Principal components (PCs) loading by maximizing the dispersion and minimizing the number of co-efficient. This tool extracts latent information and discriminates parameters with a large load. The PCs are used for varimax rotation and generate varifactors (VFs), including hypothetical, latent, and unobservable values, to gather the details (Helena et al., 2000). FA simplifies the data structure coming from PCs and decreases the contribution of low-significant observations (Bu et al., 2010).

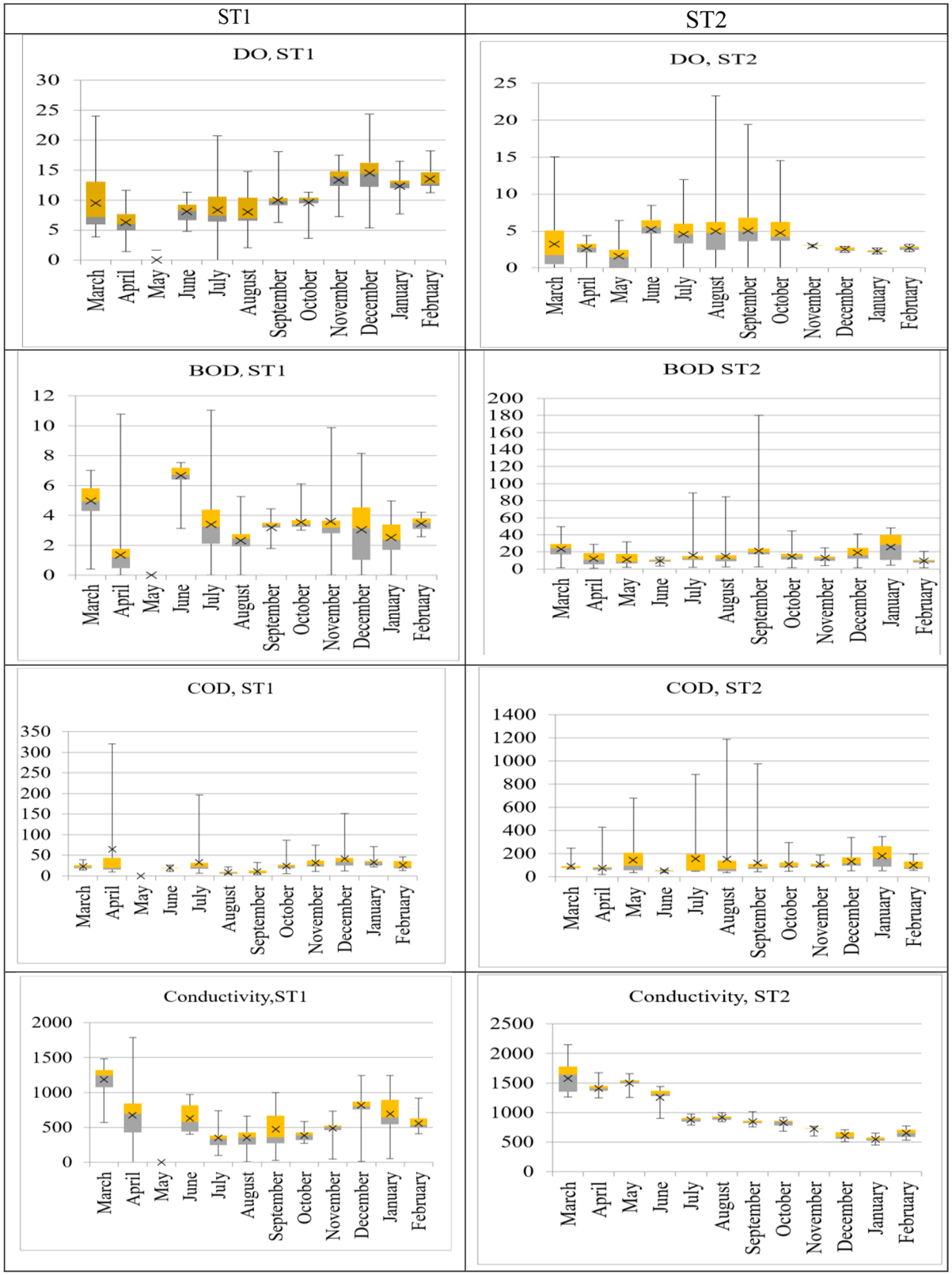
$$z_{ij} = a_{f1}f_{i1} + a_{f2}f_{i2} + \cdots + a_{fm}f_{mi} + e_{fi} \quad (2)$$

where z = observed variables, a and e are factor loading and score, e = residual term associated with error and other variations, i = sample no, m = no of factor.

For PCA/FA, water quality datasets were arranged into four main seasons: Summer, monsoon, post-monsoon, and winter. The data were standardized to classify correctly for the higher difference in data dimension (Liu et al., 2003). Standardization makes variables dimensionless to enhance their influence with different variances by eliminating the different measurement units (Singh et al., 2004). PCA was applied, and the variance, covariance, loading, and score were calculated. All the computations were performed using Excle2019, SPSS 20.0, and Origin Pro-2022.

3 Results

The water quality data for 11 parameters were collected from March 2021 to February 2022. For ST1, the dataset for May was not available. Tables 1 and 2 show the mean values of eleven water quality parameters on monthly basis Data were analyzed for spatial and temporal variations, and Box Whisker plots are shown in Figs. 2 and 3. For ST1, the mean DO varied between 4.64 and 12.8 mg/l. Whereas for ST2, it varied from 1.62 to 5.43 mg/l. BOD varied from (1.37 ± 1.34) mg/l to (6.65 ± 0.8) mg/l for ST1. A higher value was observed during the monsoon (July). Whereas for ST2, BOD was varied from (7.24 ± 6.75) mg/l to (25.98 ± 13.9) mg/l. During the monsoon, high BOD might be caused for the decaying of organic substances and vegetation in the river water from surface runoff (Gadhia et al., 2013). COD observed (8.15 ± 3.5) mg/l in August and (64.09 ± 88.7) mg/l in April at Wazirabad (ST1). At Okhla (ST2), COD fluctuated from (52.13 ± 6.1) mg/l to (158.02 ± 219.9) mg/l. The conductivity varied from (351.35 ± 99.47) µS/cm to (1181.92 ± 197.48) µS/cm at ST1, whereas the parameter varied from (548.55 ± 34.7) µS/cm to (1575.71 ± 234.65) µS/cm at ST2. The values for pH were found within the range for both stations. For ST1 mean pH varied from (7.22 ± 0.44) s.u. to (7.99 ± 0.5) s.u., and for ST2, pH varied from (6.0949 ± 0.3971) s.u. to (7.97 ± 0.2926) s.u. The median pH is around 7.5 t0 8, describing the river outfalls with wastewater-neutralizing water (Arora & Keshari, 2021). The mean temperature of river water was from (13.99 ± 1.1)°C to (33.9 ± 0.58)°C and (15.23 ± 0.96)°C to (30.06 ± 1.05)°C for ST1 and ST2, respectively. A high mean turbidity concentration was observed in monsoon (733.88 ± 489.4) NTU at ST1 and (239.85 ± 26.43) NTU in winter at ST2. NH4 varied from 0.02 to 1.2 mg/l for ST1, whereas 8.39 to 59.11 mg/l for ST2. Turbidity varied from 10.38 to 733.88 NTU for ST1. PCA/ FA is shown in Figs. 4,


ST1
ST2
DO, ST1
DO, ST2
BOD, ST1
BOD ST2
COD, ST1
COD, ST2
Conductivity,ST1
Conductivity, ST2
March
April
May
June
July
August
September
October
November
December
January
February

◂**Fig. 2** Box Whisker plots showing temporal and spatial variation (DO, BOD, COD, and Conductivity) (all parameters are in mg/l except Conductivity (µS/cm)

5, and 6. Factors and their variances are shown in Tables 3, 4, 5, and 6. At ST1, DO concentration was high in December (12.84 ± 2.96) mg/l. At ST2, DO was found high in September (4.98 ± 4.47) mg/l.

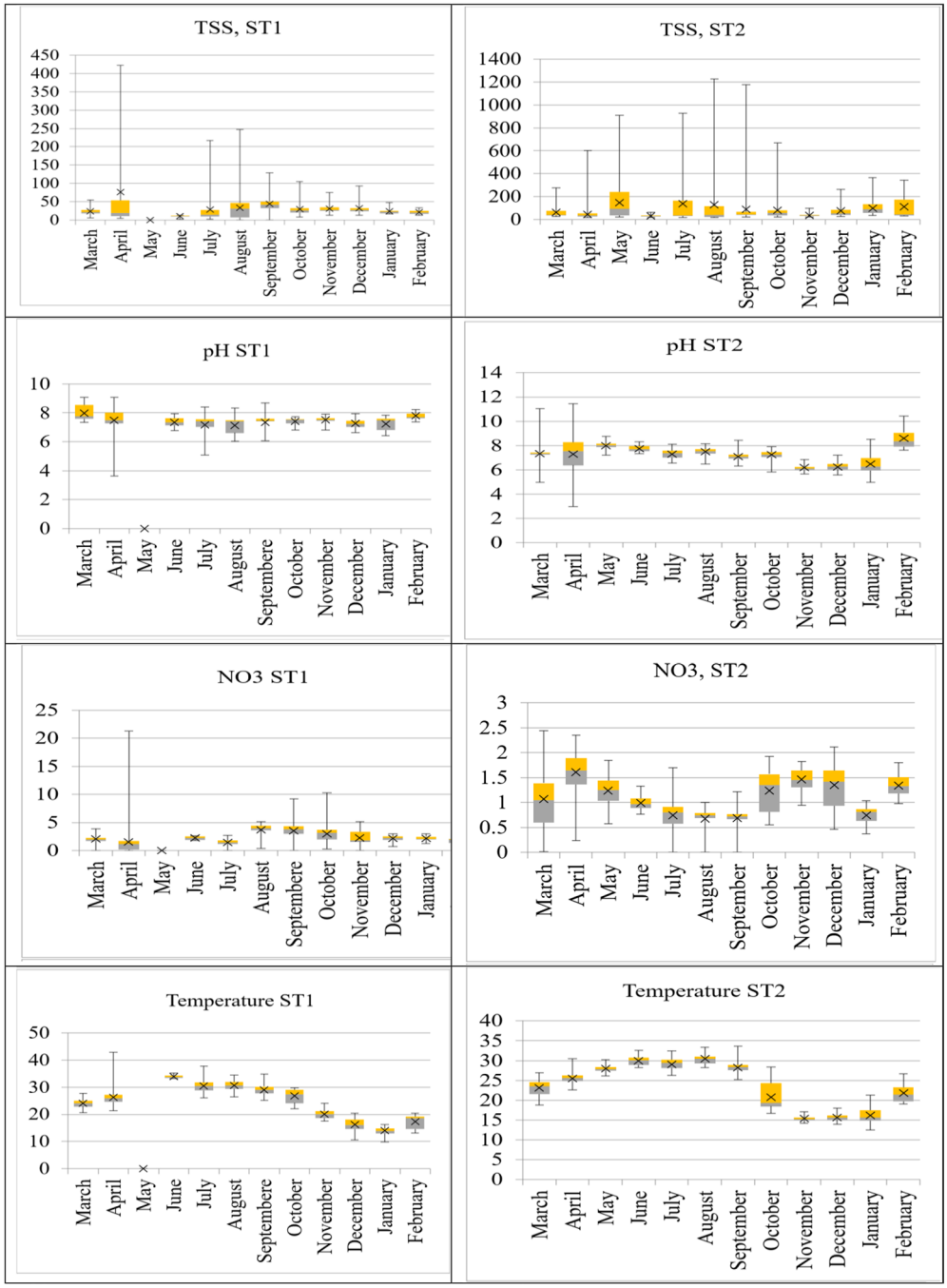


Fig. 3 Box Whisker plots for spatial/temporal variations such as TSS, NH_4, NO_3, and TOC are in mg/l, pH (s.u.), temperature (°C), and turbidity (NTU)

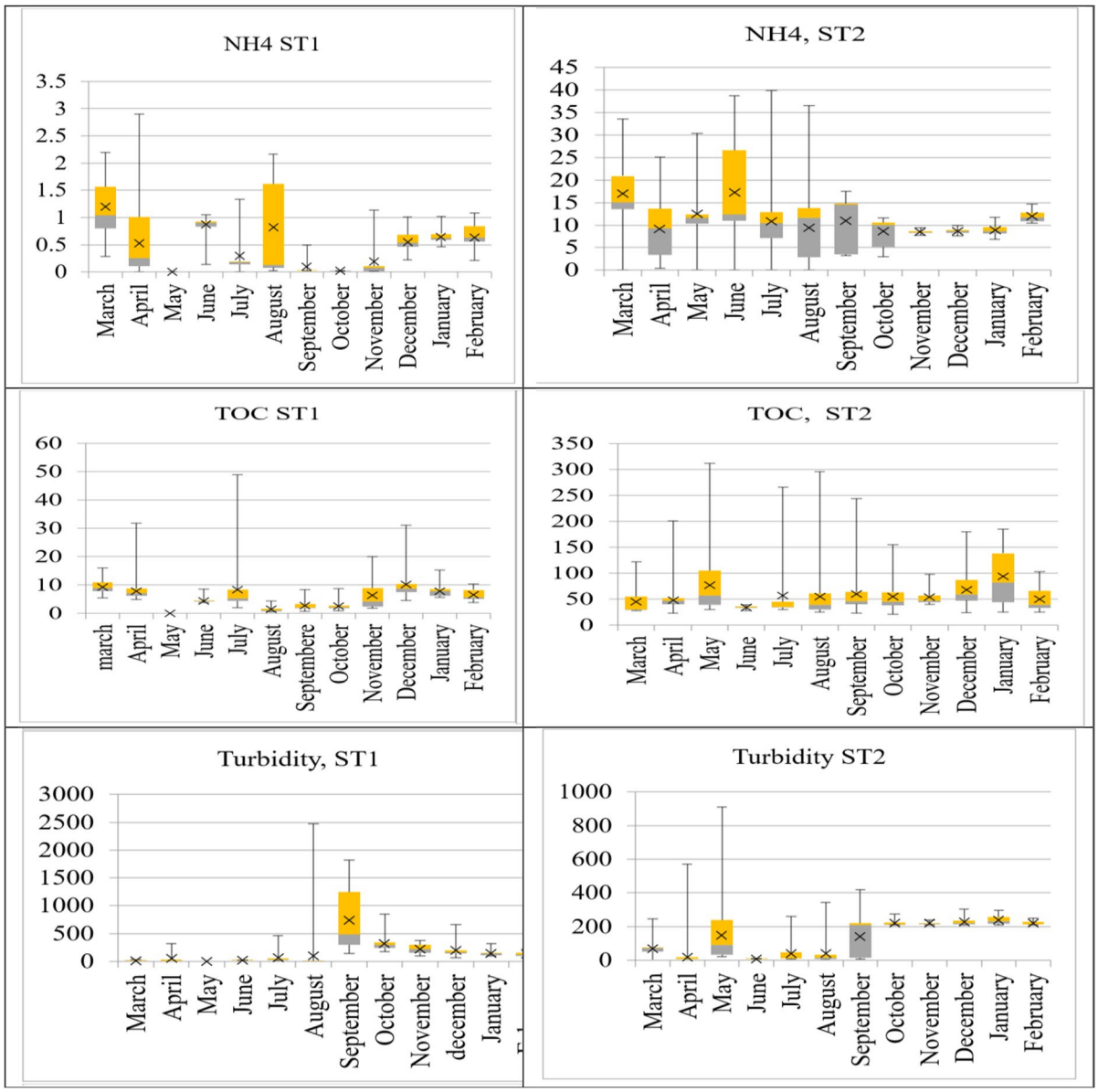


Fig. 3 (continued)

Maximum COD found during monsoon (Fig. 2). TSS is measured in fine clay, silt, microorganisms like plankton, and organic and inorganic substances. TSS value was high in lean summer flow (76.95 ± 1) mg/l and (147.5 ± 138.6) mg/l for ST1 and ST2, respectively.

For ST1, in summer, four PCs with eigenvalues > 1 explain about 84% of the whole variance in the data. The first PC counted 41% of the combined variance and appeared in positive correspondence with COD, turbidity, TSS, and TOC. The second PC that contained 22.3% of the aggregated variance and positive loading were conductivity, NO3, pH, and DO. Factor 3 had a strong negative relation with temperature. BOD had moderate loading in factor 2. For ST2, summer season, PCs were four with 80% of the aggregated variance, and PC1 had 35% of the variance. Factors were determined, and factor 1 was strongly correlated with TOC, COD, and TSS. Factor

Table 1 Statistics for the station ST1

Month	DO (mg/L)	Conductivity (µS/cm)	NO3 (mg/l)	TOC (mg/l)	TSS (mg/l)	pH	Temp (°C)	NH4 (mg/l)	Tur (NTU)	BOD (mg/l)	COD (mg/l)
March	7.87 ± 4.99	1181.92 ± 197.48	2.08 ± 0.53	9.31 ± 2.28	23.92 ± 8.8	7.99 ± 0.5	24.04 ± 1.49	1.2 ± 0.52	10.38 ± 3.85	4.96 ± 1.11	22.38 ± 5.34
April	4.64 ± 1.91	674.92 ± 342.9	1.52 ± 1.88	23.61 ± 31.36	76.95 ± 117.14	7,47 ± 0.997	26.31 ± 3.12	0.53 ± 0.62	54.65 ± 89.34	1.37 ± 1.34	64.09 ± 88.7
May	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
June	6.42 ± 1.79	626.97 ± 194.05	2.22 ± 0.28	4.45 ± 0.59	10.99 ± 1.5	7.36 ± 0.33	33.9 ± 0.58	0.87 ± 0.15	15,5 ± 3.31	6.65 ± 0.8	18.88 ± 1.6
July	6.65 ± 2.91	384.80 ± 124.28	1.47 ± 0.59	8.3 ± 8.24	29.05 ± 37.96	7.17 ± 0.69	30.55 ± 2.24	0.3 ± 0.33	61,26 ± 76.86	3.4 ± 1.88	33.05 ± 33.16
August	6.34 ± 2.65	351.35 ± 99.47	3.69 ± 1.16	1.44 ± 0.99	33.64 ± 23.6	7.14 ± 0.52	30.57 ± 1.88	0.82 ± 0.76	96.67 ± 445.1	2.32 ± 0.7	8.15 ± 3.5
September	8.34 ± 1.68	471.23 ± 225.05	3.55 ± 1.21	2.69 ± 1.4	44.62 ± 23.6	7.36 ± 0.56	28.94 ± 1.72	0.09 ± 0.14	733.88 ± 489.4	3.19 ± 0.56	10.23 ± 5.45
October	7.91 ± 1.54	380.33 ± 69.45	2.98 ± 1.16	2,56 ± 1.3	29.99 ± 15.73	7.42 ± 0.21	26.82 ± 2.53	0.02 ± 0.004	312.01 ± 115.87	3.53 ± 0.48	23.9 ± 14.74
November	11.70 ± 2.02	485.21 ± 97.42	2.30 ± 1.21	6.37 ± 4.7	31.21 ± 12.13	7,52 ± 0.17	20.12 ± 1.57	0.19 ± 0.29	219.49 ± 73.2	3.58 ± 1.88	31.22 ± 13.16
December	12.84 ± 2.96	817.54 ± 133.17	2.13 ± 0.47	10.22 ± 5.14	32.73 ± 12.14	7.26 ± 0.27	16.25 ± 1.99	0.55 ± 0.21	197.95 ± 114.6	3.02 ± 2.11	41.08 ± 27.15
January	10.70 ±	694.05 ± 212.37	2.13 ± 0.4	7.89 ± 1.9	23.73 ± 5.84	7.22 ± 0.44	13.99 ± 1.1	0.65 ± 0.07	142.4 ± 43.3	2.48 ± 1.1	32.77 ± 9.68
February	11.87 ±	558.2 ± 92.73	1.84 ± 0.39	6.49 ± 1.83	21.61 ± 5.28	7.79 ± 0.2	17.4 ± 2.37	0.64 ± 0.24	127.45 ± 38.04	3.45 ± 0.4	26 ± 9.5

Table 2 Statistics for the station ST2

Month	DO (mg/l)	Conductivity (µS/cm)	NO3 (mg/l)	TOC (mg/l)	TSS (mg/l)	pH	Temp (mg/l)	NH4 (mg/l)	Tur (NTU)	BOD (mg/l)	COD (mg/l)
March	3.88 ± 3.96	1575.1 ± 234.7	1.07 ± 0.57	47.49 ± 37.5	63.3 ± 53.6	7.3334 ± 0.25	23.37 ± 1.49	17.65 ± 8.97	68.9 ± 46.07	23.23 ± 8.12	93.02 ± 36.4
April	2.55 ± 0.76	1406.6 ± 58.6	1.5 ± 0.31	47.5 ± 13.5	43.87 ± 34.4	7.28 ± 1.3	23.56 ± 1.07	59.13 ± 6.78	18.68 ± 33.12	11.97 ± 6.98	76.23 ± 30.98
May	1.62 ± 1.72	1494.7 ± 86.9	1.24 ± 0.26	76.8 ± 45.4	147.5 ± 138.6	7.97 ± 0.29	27.92 ± 0.68	10.93 ± 2.31	47.5 ± 138.7	5.61 ± 1.76	143.44 ± 103.4
June	5.23 ± 1.94	1251.6 ± 170.2	0.79 ± 0.3	34.87 ± 3.12	30.68 ± 5.56	7.27 ± 0.14	30.06 ± 1.05	17.38 ± 10.02	8.8 ± 0.96	9.74 ± 1.88	51.13 ± 3.81
July	4.55 ± 2.18	871.87 ± 45.78	0.74 ± 0.3	93.04 ± 125.3	138.64 ± 236.83	7.47 ± 0.34	29.06 ± 1.5	10.84 ± 8.21	39.9 ± 66.6	16 ± 11.58	157.02 ± 228.5
August	4.76 ± 3.75	912.19 ± 34.4	0.68 ± 0.23	89.72 ± 127.3	128.83 ± 234.95	7.0 ± 0.27	30.4 ± 1.15	9.45 ± 6.07	38.5 ± 65.43	15.67 ± 12.28	157.02 ± 221.9
September	4.98 ± 4.47	841.3 ± 32.55	0.7 ± 0.16	71.88 ± 81.14	88.26 ± 151.8	7.01 ± 0.2	28.04 ± 1.08	10.97 ± 5.35	39.74 ± 106.6	21.09 ± 7.42	118.73 ± 99.4
October	3.15 ± 0.93	826.19 ± 48.7	1.24 ± 0.39	58.69 ± 44.9	77.69 ± 88.3	7.12 ± 0.22	27.5 ± 1.62	8.67 ± 2.8	220.9 ± 14.6	14.42 ± 5.3	110.35 ± 50.2
November	2.03 ± 0.09	724.62 ± 17.16	1.1 ± 0.14	39.87 ± 8.68	34.5 ± 7.9	6.16 ± 0.17	20.14 ± 0.4	11.05 ± 0.5	18.9 ± 6.7	12.75 ± 4.43	108.57 ± 25.9
December	2.34 ± 0.11	606.8 ± 54.86	1.5 ± 0.19	61.69 ± 30.4	75.8 ± 59.2	6.74 ± 0.63	16.86 ± 1.5	8.58 ± 0.83	226.01 ± 14.13	18.81 ± 9.73	133.46 ± 52.25
January	2.43 ± ± 0.26	548.55 ± 34.7	1.05 ± 0.38	79.48 ± 31.1	96.8 ± 50.3	6.09 ± 0.8	15.23 ± 0.96	8.39 ± 0.5	239.85 ± 26.43	25.98 ± 13.95	182.54 ± 94.24
February	3.27 ± 0.32	646.79 ± 60.1	0.87 ± 0.24	74.29 ± 57,9	110.3 ± 104.8	7.18 ± 0.7	17.96 ± 1.67	9.8712 ± 0.92	0.67 ± 1.18	9.17 ± 2.98	100.56 ± 41.93

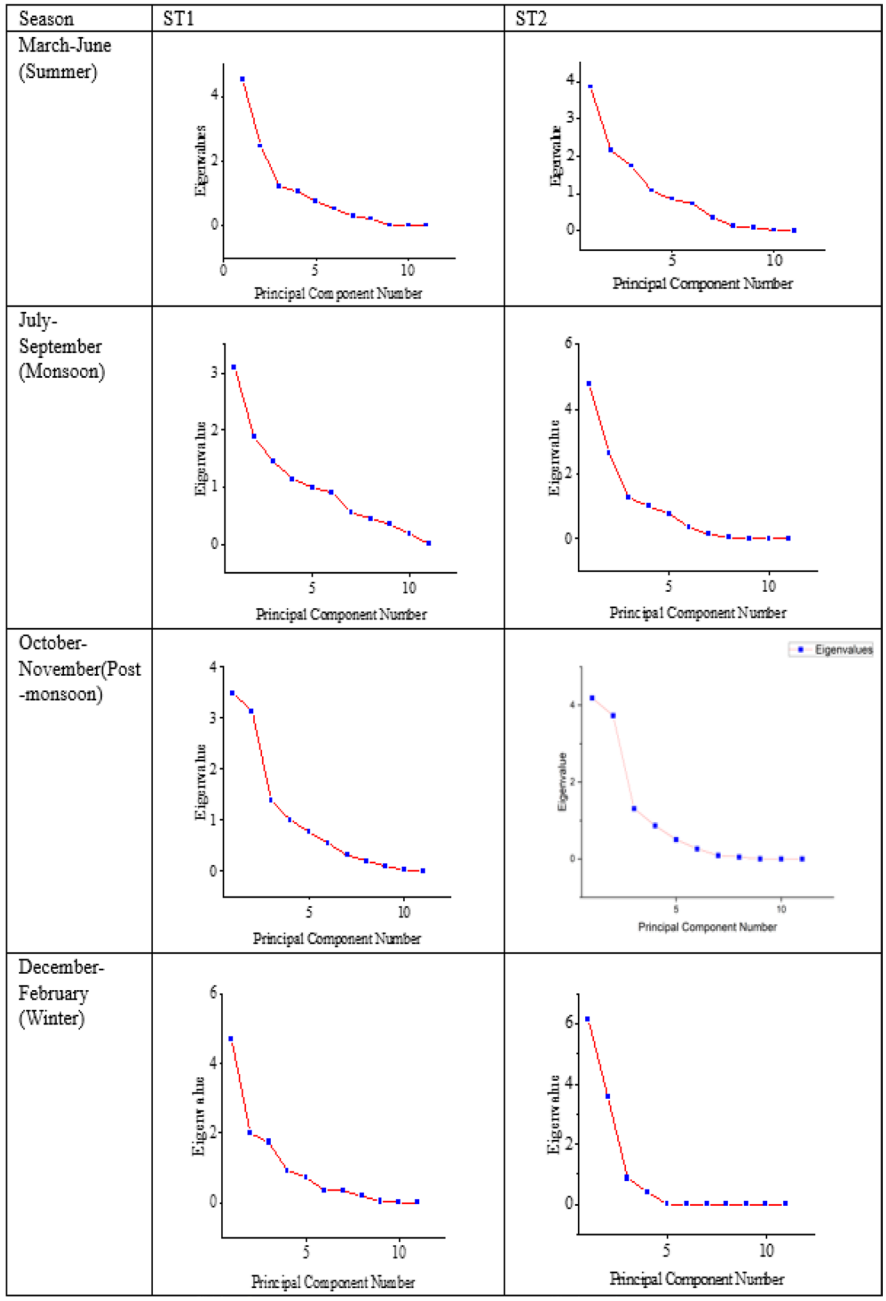


Fig. 4 Scree plots for determining principal components of four seasons

2, factor 3, and factor 4 were correlated with conductivity, NH4, and NO3, respectively. Factor 3 was strongly negatively related to pH. For monsoon, PCs were 4 for the ST1 with a variance of 68%, with PC1 28.2% variance. TOC, COD, and TSS were strongly correlated with PC1. Factor 1 was strongly related to TOC and COD, moderate loading with TSS, and very low negative turbidity loading. For ST2, 3 PCs were obtained for the monsoon period with a 78.8% whole variance. For monsoon, TOC and COD were strongly related to factor 1 at ST1. For factor 2 and factor 3 NH4 and turbidity were strongly related. At ST2, for factor 1 TSS, COD, and TOC were positively related while NO3 was negatively strongly related. For factor 2, pH, conductivity, and temperature were strongly related and for factor 3, BOD was strongly related.

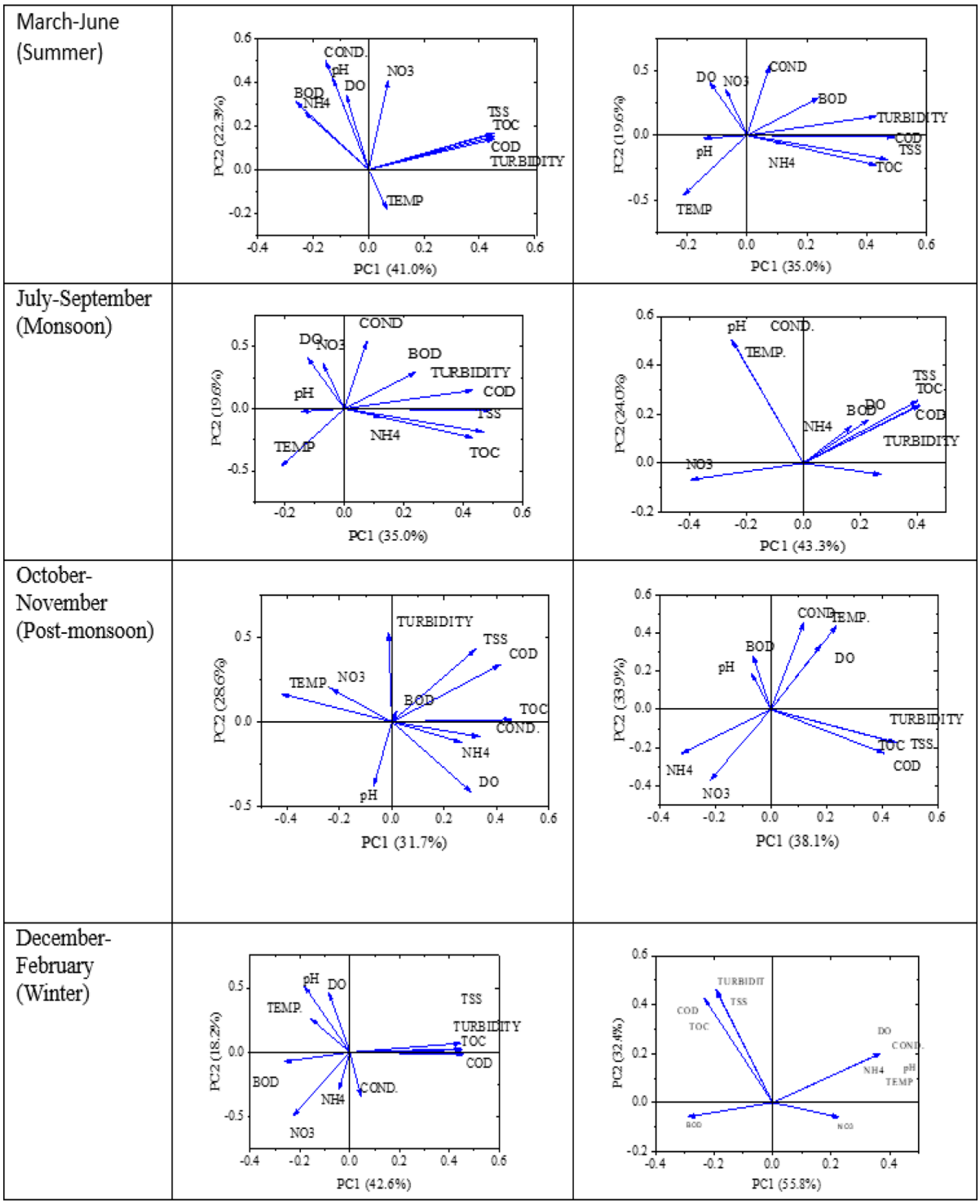


Fig. 5 Loading diagram for principal components (PCs)

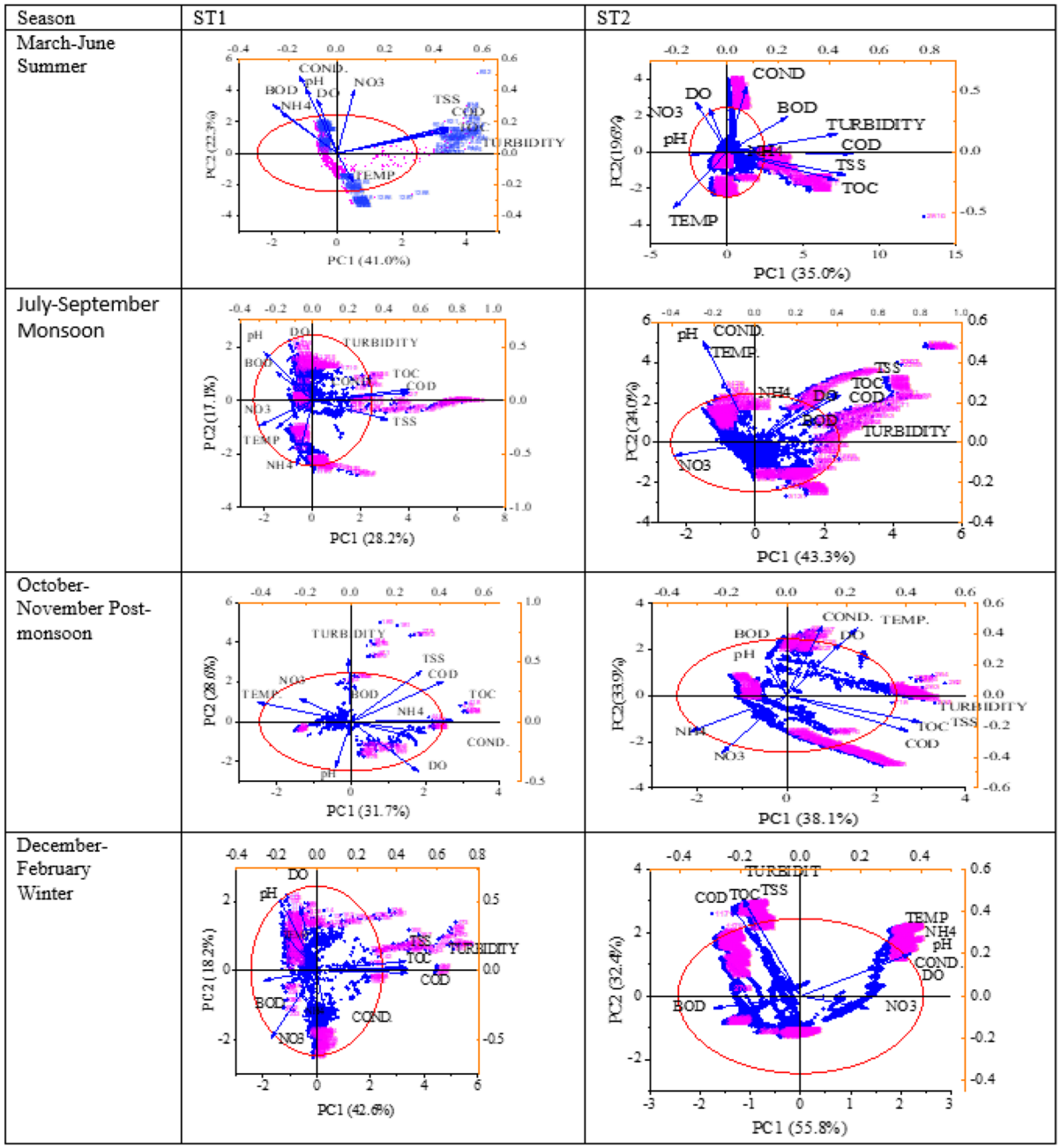


Fig. 6 Biplot for principal components (PCs)

For post-monsoon, ST1 had four PCS with a total variance of 82%. Factor 1 was positively related to Turbidity, COD, and TSS. Factor 2 was strongly positively related to conductivity and moderately related to NH4. Factor 3 was strongly negatively related to NO3 and factor 4 was strongly positively related to BOD. AT ST2, factor 1 was strongly related to turbidity, TOC, COD, and TSS. Factor 2 was strongly related to conductivity and temperature while moderately related to DO. Whereas, factor 3 was strongly related to pH. For winter, at ST1, factor 1 had a positive strong correlation with turbidity, TOC, COD,

Table 3 Factors for water quality parameters at two stations–summer season (bold values show strong correlation)

ST1					ST2			
Variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3	Factor 4
DO	−0.03	0.203	−0.172	**0.8932**	−0.341	0.585	−0.09	0.038
Turbidity	**0.988**	−0.038	−0.068	−0.05	0.73	0.485	0.1197	−0.377
TOC	**0.992**	−0.005	−0.07	−0.022	**0.93**	−0.277	0.0737	0.111
NO3	0.302	**0.774**	−0.055	0.0688	−0.103	0.209	0.0681	**0.909**
BOD	−0.42	0.749	0.002	0.1775	0.489	0.433	−0.269	0.212
NH4	−0.311	0.248	0.571	0.2082	0.097	−0.101	**0.9197**	0.162
COD	**0.988**	−0.005	−0.096	−0.037	**0.976**	0.101	0.0829	0.04
pH	−0.036	0.087	0.466	**0.796**	−0.074	−0.135	**−0.913**	0.092
Cond	−0.072	**0.759**	0.478	0.1635	−0.015	**0.841**	0.0867	0.04
Temp	−0.02	0.027	**−0.851**	0.0636	−0.326	−0.673	−0.059	−0.281
TSS	**0.993**	−0.018	−0.047	−0.014	**0.908**	−0.044	0.1179	−0.321

Table 4 Factors for water quality parameters at two stations– monsoon season (bold values show strong correlation)

ST1					ST2		
Variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3
DO	−0.019	−0.2454	0.4236	0.4848	0.4001	−0.0169	0.7237
Turbidity	−0.004	−0.1918	**0.8903**	−0.043	0.4687	−0.3441	0.1549
TOC	**0.9644**	0.0021	0.0229	−0.014	**0.9746**	−0.0544	0.1144
NO3	−0.627	0.3781	0.4433	−0.018	**−0.888**	0.2755	0.13
BOD	−0.179	−0.3799	−0.0893	0.6289	0.2101	−0.0084	**0.8866**
NH4	−0.26	**0.805**	−0.1565	0.0435	−0.1087	−0.0177	0.2724
COD	**0.963**	0.0419	0.0189	0.0051	**0.9777**	−0.0602	0.1185
pH	−0.402	−0.6727	0.1362	0.0361	−0.1092	**0.9714**	−0.0044
Cond	0.1165	0.0298	0.0152	0.2819	−0.1084	**0.9843**	−0.0232
Temp	−0.251	0.3781	−0.1271	0.7202	−0.1085	**0.9843**	−0.0234
TSS	0.6285	0.4729	0.4796	−0.163	**0.97**	−0.0231	0.1029

Table 5 Factors for water quality parameters at two stations– post-monsoon season (bold values show strong correlation)

ST1					ST2		
Variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 1	Factor 2	Factor 3
DO	−0.426	0.5613	0.6057	0.1364	0.0377	**0.7544**	0.0698
Turbidity	**0.8242**	−0.233	−0.396	0.0571	**0.9752**	0.1149	−0.1
TOC	0.3962	0.3312	0.7243	0.3421	**0.9611**	−0.033	−0.029
NO3	0.0526	−0.025	**−0.862**	0.3269	−0.032	−0.87	0.1333
BOD	0.0575	−0.064	−0.091	**0.7544**	−0.358	0.4257	0.1107
NH4	0.0274	0.7428	0.0651	−0.451	−0.282	**−0.77**	0.4656
COD	**0.8752**	0.3417	0.2582	0.1317	**0.9611**	−0.033	−0.029
pH	−0.7	0.0603	0.1203	0.4261	−0.094	0.1281	**0.9688**
Cond	0.1037	**0.8099**	0.0535	−0.039	−0.097	**0.8455**	0.4758
Temp	−0.034	−0.738	−0.366	−0.429	0.0752	**0.9568**	0.1938
TSS	**0.9442**	0.1164	0.154	0.1713	**0.972**	0.1167	−0.098

and TSS. However, at ST2, factor 1 had a positive strong relation with DO, NH4, pH, conductivity, and temperature. Whereas, factor 2 had a relation with turbidity, TOC, COD, and TSS.

Table 6 Factors for water quality parameters at two stations– winter season (bold values show strong correlation)

ST1				ST2	
Variables	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2
DO	−0.0759	**0.783517**	−0.18294	**0.988745**	−0.06384
TURBIDITY	**0.976323**	−0.01286	0.05929	−0.0445	**0.990699**
TOC	**0.985584**	−0.01003	0.080926	−0.16006	**0.977882**
NO3	−0.52831	−0.20446	0.692523	0.456462	−0.35217
BOD	−0.53339	0.310666	0.351375	−0.69496	0.221442
NH4	−0.17709	−0.573	0.016162	**0.98873**	−0.0637
COD	**0.977242**	−0.12436	0.042429	−0.16006	**0.977882**
pH	−0.34638	0.388568	−0.67418	**0.988694**	−0.06498
COND	0.125873	0.207543	**0.859875**	**0.988737**	−0.06371
TEMP	−0.23439	**0.781761**	0.180959	**0.988739**	−0.06368
TSS	**0.982774**	0.069416	0.05171	-0.03966	**0.990255**

4 Discussions

The data assessment observed that DO concentration varied sharply among the two stations, revealing that substantial oxygen-consuming pollutants were discharged through the sixteen main drains (Fig. 1) (Mandal et al., 2010). At ST1, DO concentration was found high in December. It might be due to the water's low temperature and high inflow volume, as DO is inversely related to temperature (Water Science School, 2018). Whereas, at ST2, DO concentration was found high in September due to the integrated effect of high wind speed and freshwater mixing due to rainfall (Das et al., 1997). DO is an indicative parameter for the fitness of an aquatic ecosystem, the metabolism, and the respiration of aquatic organisms (Gadhia et al., 2013). At Wazirabad, the minimum DO was adequate to sustain lives. However, at the ST2, the minimum DO was observed as nil for most of the year (Arora & Keshari, 2021). High BOD and COD were observed during some periods of monsoon. The monsoon brought runoff from agricultural fields, raising the nutrient concentration (Gadhia et al., 2013). The high COD load is caused by surface runoff, indicating organic pollution caused by non-point sources. (Joseph and Srivastava (1993) concluded that higher BOD observed during the monsoon period indicated high demands of oxygen required for life process support. Clark, 1986, also stated that higher organic load experienced in monsoon caused higher BOD levels. Due to continuous rainfall, soil erosion and surface runoff with non-point pollutants are attributed to additional organic waste load in the river water and increased BOD level (Odokuma & Okpokwasili, 1997). The large value of BOD and COD indicated the presence of organic–inorganic chemicals (Arora & Keshari, 2021). Furthermore, it clearly stated the presence of high oxygen-demanding pollutants from both fixed and diffused sources. The chemical oxygen demand was high in the monsoon. Since BOD and COD are highly related, the causes for BOD and COD being high during the monsoon accounted for unchanged. Industrial effluents carried high COD resulting in high values of oxygen-demanding chemicals in river water (Gadhia et al., 2013). The summer season showed a higher conductivity range might occur due to the lower flow available in the river during summer (Gadhia et al., 2013). The conductivity value decreased during the monsoon at both stations. It might be a high volume of water from the upstream end due to rainfall. In the monsoon, river water gets diluted due to the increasing volume of water (Izonfuo & Bariweni, 2010). Higher values of TSS at ST2 indicated wastewater outfall with substances increasing TDS. During the monsoon, a low TSS value was observed for both points. TSS can be changed by changing the pH, which causes precipitated solutes and affects solubility (Gadhia et al., 2013). pH was observed higher in the summer and low in the monsoon. Variations in pH values are attributed to factors like dilution of water (Rajasegar, 2003), reduction of temperature, decaying of organic substances, and degradation of carbon dioxide (CO2) during photosynthesis through reduction of bicarbonate (Izonfuo & Bariweni, 2010). The higher conductivity, BOD, COD, and low DO express the outfalling point

sources and diffused sources contributing pollutants into the river stretch (Arora & Keshari, 2021). Two locations did not show significant temperature variation revealing the outfalling drains did not contribute to high-temperature wastewater (Arora & Keshari, 2021). Maximum turbidity was found in the pre-monsoon period (Fig. 3). Turbidity in water shows the presence of suspended matter like silts, clay, plankton, and other microorganisms, organic and inorganic substances. NO3 concentration was high during monsoon at ST1 and could be due to organic substances discharged from the surroundings (Das et al., 1997). NH4 level was low during the same period. The cause may be the oxidation of ammonia to nitrite and nitrate (Rajasegar, 2003). Figure 3 shows that the availability of sufficient flow due to rainfall reduces the ammonia concentration in river water (Arora & Keshari, 2021). A high variation of ammonia concentration was observed between the two locations, and it concluded that the river got discharged with a high volume of nutrient-rich wastewater. TOC concentration was low in the monsoon period due to sufficient flow and high in the summer season at Wazirabad. A higher concentration of TOC was observed at Okhla during the monsoon, attributed to the presence of the organic substance in surface runoff. For PCA/FA total dataset has been divided into four seasons for both locations: summer, monsoon, post-monsoon, and winter. The PCA reduces the dimension of data keeping information unchanged and identifying the eigenvector and eigenvalue (Arora & Keshari, 2021). Scree plots (Fig. 4) were used to determine the principal components for each dataset to perceive the data structure (Vega et al., 1998). The PCs were selected from scree plots with eigenvalues of more than one. Loadings of PCs and actual variables are shown in Fig. 5. The biplot provides correlation and impact of different parameters with PCs shown in Fig. 6 for multiple locations. A new set of factors were defined by rotating the PC's axis. New factors introduced involving primarily a subset of original variables without overlapping each other and independent groups shown in Fig. 7.

FA and PCA were used for the standardized data for correlating the structural designs of the water parameter and identifying the parameters with influence on factors. The first PC showed positive correspondence with COD, turbidity, TSS, and TOC for ST1. The factor loaded with TSS indicates that the runoff originated from a field of high solids and waste (Singh et al., 2004). The scree plot was used to select the no of PCs with an eigenvalue greater than 1 (Vega et al., 1998). FA further examined the values of PCs and identified the factors that participated more clearly. Liu et al. (2003) explained the factor loading as strong (>0.75), moderate ($0.75-0.50$), and weak ($0.50-0.30$), correlating to absolute loading. Four factors identified with 84% of the variance, TOC, Turbidity, COD, and TSS, were strong loading with factor 1. The excessive loading of these parameters might be due to low flow, erosive process, presence of microorganisms, and huge suspended organic and inorganic substances from upstream (Fan et al., 2010). NO_3 and conductivity were correlated with factor 2, and Factor 4 strongly correlated with DO and pH. The strong relationship between NO3 and conductivity might be due to high agricultural runoff (Ogwueleka, 2015). Factor 3 had a strong negative correlation with temperature. BOD had moderate loading in factor 2, indicating an excessive loading of dissolved organic matter added from urban runoff carrying nutrients with high oxygen demand (Pati et al., 2014). For ST2, summer season, PCs were four with 80% of the aggregated variance, and PC1 had 35% of the variance. Factor 1 strongly correlated with TOC, COD, and TSS. These factors loading indicated that domestic wastewater, industrial discharges, and agricultural runoff had strongly affected the water quality.

For monsoon, TOC, COD, and TSS were strongly related to PC1 at ST1. It revealed that compared to the summer season, in monsoon, sufficient water flow was available in the location, which reduced the turbidity and TSS. However, organic and inorganic substances with a high oxygen demand were attributed. Factor 2 and factor 3 were strongly loaded with NH4 and turbidity, respectively. Factor 4 was not strongly correlated with any variables, and moderate loading with BOD indicated the presence of moderate oxygen-demanding organic substances PC1 with 43.3% variability positively correlated with TSS, TOC, and COD, indicating domestic, industrial, and surface runoff with a high volume of organic and inorganic substances, microorganisms, and fine materials (Akbal et al., 2011).

For post-monsoon, ST1 had four PCs with 82% of the total variance. Factor 1 loading revealed low flow conditions with high turbidity, TSS, and COD. Whereas, factor 2 was strongly loaded with conductivity and moderate loading with NH_4. The reasons

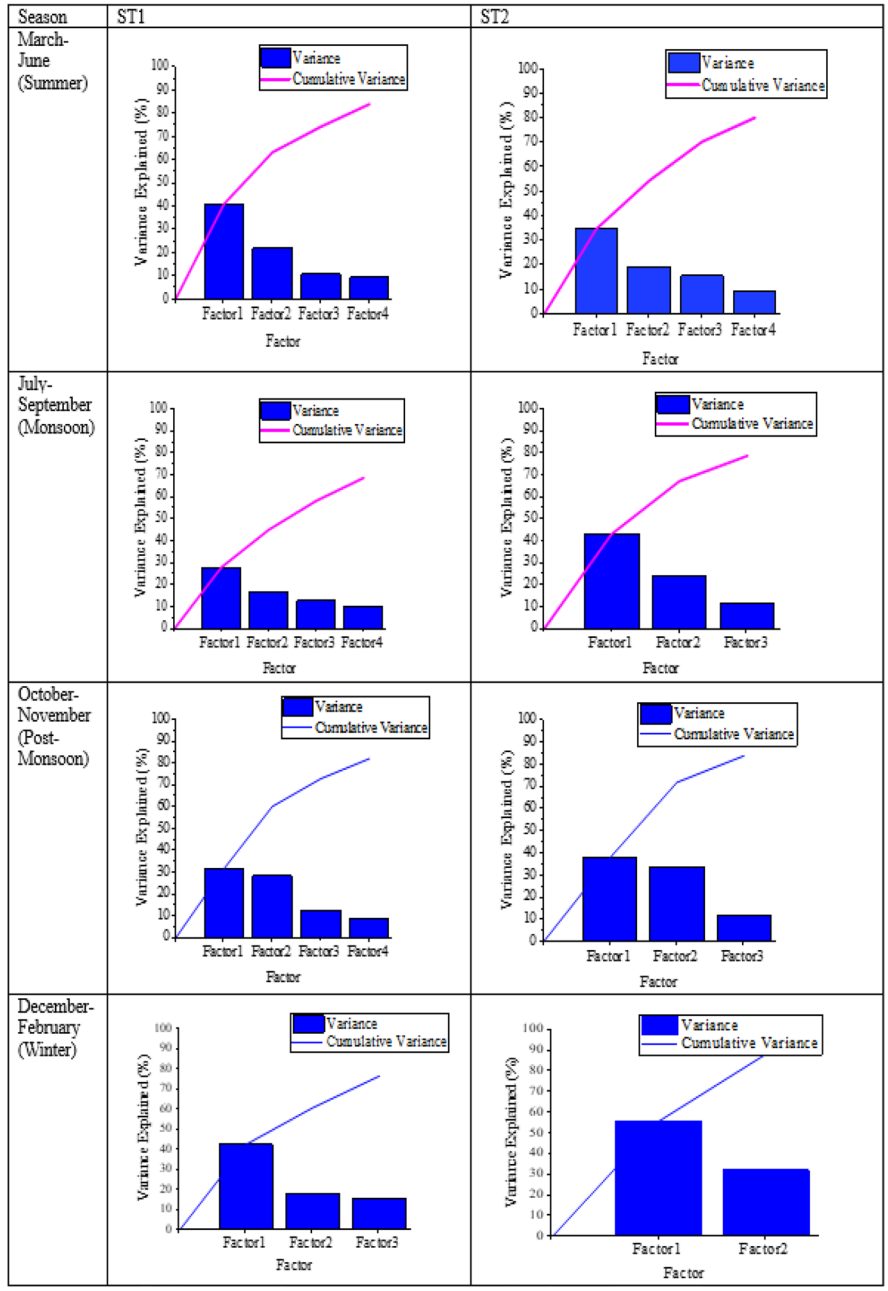


Fig. 7 Variances for factors

could be upstream flow with agricultural runoff with high nutrients, Factor 3 had strong negative loading of NO_3, and Factor 4 had strong positive loading of BOD. For ST2, three PCs with 83.9% variance and DO had strongly positive loading in factor 2. For winter, ST1. 76.5% of total variance with 46.2%variance in PC1. For ST2, PCs were two with 86.2% of the total variance, and PC1 had the 55% of the variance. For Factor 1, intense positive loading was Turbidity, TOC, COD, and TSS, indicating 55% of total variance with anthropogenic sources of pollution. Factor 2 showed a high positive relation with DO, temperature, and conductivity, explaining the physical characteristics of the river. This element illustrated that DO in water depends on temperature (Bu et al., 2010). However, significant loading of BOD and COD

revealed that partially treated and untreated wastewater discharged into the river contained pollutants.

In contrast, the strong loading of TKN and NH4 indicate the transfer of excess nutrient from industrial effluent and agricultural runoff (Singh et al., 2020). Negative loading of pH and DO also indicate contamination with industrial effluents forming anaerobic conditions with high pH and low oxygen. The wastewater carrying industrial effluent and domestic sewage is outfalling throughout the length between ST1 and ST2. The factor analysis of the Yamuna River reveals anthropogenic pollution sources and a high volume of organic and inorganic substances demanding a large amount of oxygen, hence undergoing an anaerobic fermentation process forming ammonia and other organic acids(Singh et al., 2004).

5 Conclusions

This study used PCA/FA and Box Whisker plots to predict seasonal variations of river quality parameters upstream and downstream of the Yamuna River Delhi stretch. The methods were used to reduce the data set and assisted in extracting and establishing the significant constituents/ sources accountable for the changeability of water quality parameters at two monitoring locations. The spatiotemporal variations revealed that the downstream water quality is severely poor throughout the year due to the point and diffused sources of pollution. Although there was sufficient flow in the monsoon, the water quality was not maintained to the standard due to surface runoff with excessive oxygen-demanding organic and inorganic substances. In contrast, upstream water quality is less polluted. However, high turbidity, COD, NO_3, and conductivity described the river flowing with nutrients, high microorganisms, and inorganic substances. The results obtained from factor analyses supported the box whisker plots. The principal component analysis was applied for four seasons to observe the correlation and variability among different parameters. The varifactors from factor analysis specified that upstream water is less polluted with organic substances than downstream water. Downstream water is polluted with organic and inorganic substances from drains (untreated/ partially treated domestic and industrial wastewater) and monsoon surface runoff with highly demanding oxygen substances. The study emphasizes that reducing pollutants from different sources can improve this river's reach and management of the ecosystem. Also, regular monitoring systems are required to sustain the aquatic ecosystem.

Acknowledgements The authors are thankful to Real-Time Water Quality Monitoring Stations at Wazirabad and Okhla, Delhi Pollution Control Committee for providing data. The authors are grateful to the reviewers for giving their valuable time and suggestions to improve the quality of the study.

Author Contribution NV has completed the research work under the supervision of GS and NA. The manuscript was written by NV and critically reviewed by GS and NA. All authors read and approved the manuscript.

Data Availability The dataset analyzed during the present study is available in RTWQMS, Delhi.

Declarations

Conflict of Interest The authors declare no competing interests.

References

Achieng', A. O. et al. (2017). Assessment of water quality using multivariate techniques in River Sosiani, Kenya. *Environmental Monitoring and Assessment*, *189*(6). https://doi.org/10.1007/s10661-017-5992-5.

Akbal, F., et al. (2011). Multivariate Statistical techniques for the assessment of surface water quality at the mid-Black Sea Coast of Turkey. *Water, Air, & Soil Pollution, 216*(1), 21–37. https://doi.org/10.1007/s11270-010-0511-0

Arora, S., & Keshari, A. K. (2021). Pattern recognition of water quality variance in Yamuna River (India) using hierarchical agglomerative cluster and principal component analyses. *Environmental Monitoring and Assessment*, *193*(8). https://doi.org/10.1007/s10661-021-09318-1.

Atakoglu, O. O., & Yalcin, F. (2022). Evaluation of the surface water and sediment quality in the Duger basin (Burdur, Turkey) using multivariate statistical analyses and identification of heavy metals. *Environmental Monitoring and Assessment, 194*(7), 484. https://doi.org/10.1007/s10661-022-10147-z

Avtar, R., et al. (2013). Hydrogeochemical assessment of groundwater quality of Bundelkhand, India Using Statistical Approach. *Water Quality, Exposure and Health, 5*(3), 105–115. https://doi.org/10.1007/s12403-013-0094-2

Bonansea, M., et al. (2018). Multivariate statistical analysis for estimating surface water quality in reservoirs. *International Journal of Hydrology Science and Technology, 8*(1), 52–68. https://doi.org/10.1504/IJHST.2018.088675

Bu, H., et al. (2010). Water quality assessment of the Jinshui River (China) using multivariate statistical techniques. *Environmental Earth Sciences, 60*(8), 1631–1639. https://doi.org/10.1007/s12665-009-0297-9

Calijuri, M. L., et al. (2012). Evaluation of the influence of natural and antrhopogenic processes on water quality in Karstic Region. *Water, Air, & Soil Pollution, 223*(5), 2157–2168. https://doi.org/10.1007/s11270-011-1012-5

Central Pollution Control Board. (2012). *River Yamuna – Waste water management plan in Delhi*. Central Pollution Control Board (Ministry of Environment and Forests, Govt. of India).

Clark, R. B. (1986). *Marine pollution* (3rd ed., pp. 433–443). Oxford Science Press Publication, Clarendon Press.

CPCB. (2006). Water quality status of Yamuna River, Assessment and development of River Basin. pp. 1–115. Available at: www.cpcb.nic.in.

Das, J., Das, S. N., & Sahoo, R. K. (1997). Semidiurnal variation of some physico-chemical parameters in the Mahanadi estuary, east coast of India. *Indian Indian Journal of Geo-Marine Sciences, 26*, 323–326.

Dehghanzadeh, R., et al. (2015). Integrated Assessment of spatial and temporal variations of groundwater quality in the Eastern Area of Urmia Salt Lake Basin Using multivariate statistical analysis. *Water Resources Management, 29*(4), 1351–1364. https://doi.org/10.1007/s11269-014-0877-7

Fan, X., et al. (2010). Assessment of river water quality in Pearl River Delta using multivariate statistical techniques. *Procedia Environmental Sciences, 2*(5), 1220–1234. https://doi.org/10.1016/j.proenv.2010.10.133

Gadhia, M., Surana, R., & Ansari, E. (2013). Seasonal variations in physico-chemical characterstics of Tapi Estuary in Hazira Industrial Area. *Our Nature, 10*(1), 249–257. https://doi.org/10.3126/on.v10i1.7811

Githaiga, K. B., et al. (2021). Water quality assessment, multivariate analysis and human health risks of heavy metals in eight major lakes in Kenya. *Journal of Environmental Management, 297*(July), 113410. https://doi.org/10.1016/j.jenvman.2021.113410

Helena, B., et al. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research, 34*(3), 807–816. https://doi.org/10.1016/S0043-1354(99)00225-0

Hinge, G., et al. (2022). Integrated groundwater quality analysis using Water Quality Index, GIS and multivariate technique: A case study of Guwahati City. *Environmental Earth Sciences, 81*(16), 412. https://doi.org/10.1007/s12665-022-10544-0

Izonfuo, L., & Bariweni, A. (2010). The effect of urban runoff water and human activities on some physico- chemical parameters of the Epie Creek In The Niger Delta. *Journal of Applied Sciences and Environmental Management*, *5*(1). https://doi.org/10.4314/jasem.v5i1.54941.

Joseph, K. O., & Srivastava, J. P. (1993). Heavy metal load in edible oyster, Crassostrea madrasensis(Preston) from the Ennore Estuary in Madras. *Journal of Environmental Biology, 14*(2), 121–127.

Khullar, S., & Singh, N. (2022). Water quality assessment of a river using deep learning Bi-LSTM methodology: Forecasting and validation. *Environmental Science and Pollution Research, 29*(9), 12875–12889. https://doi.org/10.1007/s11356-021-13875-w

Koklu, R., Sengorur, B., & Topal, B. (2010). Water quality assessment using multivariate statistical methods-a case study: Melen river system (Turkey). *Water Resources Management, 24*(5), 959–978. https://doi.org/10.1007/s11269-009-9481-7

Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment, 313*(1–3), 77–89. https://doi.org/10.1016/S0048-9697(02)00683-6

Mandal, P., Upadhyay, R., & Hasan, A. (2010). Seasonal and spatial variation of Yamuna River water quality in Delhi, India. *Environmental Monitoring and Assessment, 170*(1–4), 661–670. https://doi.org/10.1007/s10661-009-1265-2

Odokuma, L. O., & Okpokwasili, G. C. (1997). Seasonal influences of the organic pollution monitoring of the New Calabar River, Nigeria. *Environmental Monitoring and Assessment, 45*(1), 43–56. https://doi.org/10.1023/A:1005709709031

Ogwueleka, T. C. (2015). Use of multivariate statistical techniques for the evaluation of temporal and spatial variations in water quality of the Kaduna River, Nigeria. *Environmental Monitoring and Assessment*, *187*(3). https://doi.org/10.1007/s10661-015-4354-4.

Pati, S., et al. (2014). Assessment of water quality using multivariate statistical techniques in the coastal region of Visakhapatnam, India. *Environmental Monitoring and Assessment, 186*(10), 6385–6402. https://doi.org/10.1007/s10661-014-3862-y

Priya, K. L., & Arulraj, G. P. (2011). A correlation-regression model for the physicochemical parameters of the groundwater in Coimbatore city, India. *Environmental Technology, 32*(7), 731–738. https://doi.org/10.1080/09593330.2010.510852

Rajasegar, M. (2003). Physico-chemical characteristics of the Vellar estuary in relation to shrimp farming. *Journal of environmental biology*, *24*(1), 95—101. http://europepmc.org/abstract/MED/12974418.

Rana, A., et al. (2012). Trend analysis for rainfall in Delhi and Mumbai, India. *Climate Dynamics, 38*(1–2), 45–56. https://doi.org/10.1007/s00382-011-1083-4

Setia, R., et al. (2021). Hydrochemical evaluation of surface water quality of Sutlej river using multi-indices, multivariate statistics and GIS. *Environmental Earth Sciences, 80*(17), 565. https://doi.org/10.1007/s12665-021-09875-1

Sharma, D., & Kansal, A. (2011). Water quality analysis of River Yamuna using water quality index in the national capital territory, India (2000–2009). *Applied Water Science*, *1*(3–4), 147–157. https://doi.org/10.1007/s13201-011-0011-4

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling and Software, 22*(4), 464–475. https://doi.org/10.1016/j.envsoft.2006.02.001

Singh, K. P., et al. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) - A case study. *Water Research, 38*(18), 3980–3992. https://doi.org/10.1016/j.watres.2004.06.011

Singh, G. et al. (2020). Assessment of spatial and temporal variations in water quality by the application of multivariate statistical methods in the Kali River, Uttar Pradesh,

India. *Environmental Monitoring and Assessment*, *192*(6). https://doi.org/10.1007/s10661-020-08307-0.

Upadhyay, R., et al. (2011). Managing water quality of River Yamuna in NCR Delhi. *Physics and Chemistry of the Earth, 36*(9–11), 372–378. https://doi.org/10.1016/j.pce.2010.03.018

Vega, M., et al. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research, 32*(12), 3581–3592. https://doi.org/10.1016/S0043-1354(98)00138-9

Venkatramanan, S., et al. (2014). Assessment of river water quality via environmentric multivariate statistical tools and water quality index: A case study of Nakdong River Basin, Korea. *Carpathian Journal of Earth and Environmental Sciences, 9*(2), 125–132.

Vieira, J. S., et al. (2012). Surface water quality assessment of Lis River Using multivariate statistical methods. *Water, Air, & Soil Pollution, 223*(9), 5549–5561. https://doi.org/10.1007/s11270-012-1267-5

Wang, Y., et al. (2013). Assessment of surface water quality via multivariate statistical techniques: A case study of the Songhua River Harbin region, China. *Journal of Hydro-Environment Research, 7*(1), 30–40. https://doi.org/10.1016/j.jher.2012.10.003

Water Science School. (2018). *Dissolved oxygen and water*. U.S. Geological Survey's (USGS).

Yilma, M., et al. (2019). Assessment and interpretation of river water quality in Little Akaki River using multivariate statistical techniques. *International Journal of Environmental Science and Technology, 16*(7), 3707–3720. https://doi.org/10.1007/s13762-018-2000-8

Zhang, Y., et al. (2009). Water quality assessment and source identification of Daliao river basin using multivariate statistical methods. *Environmental Monitoring and Assessment, 152*(1–4), 105–121. https://doi.org/10.1007/s10661-008-0300-z

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Attention-free based dual-encoder mechanism for Aspect-based Multimodal Sentiment Recognition

Pankaj Gupta
Advanced Photonics Simulator Lab
Department of Applied Physics
Delhi Technological University
Delhi-110042, India
pankajgupta.dtu@gmail.com

Ananya Pandey
Biometric Research Laboratory, Department of Information Technology,
Delhi Technological University
Delhi-110042, India
ananyaphdit08@gmail.com

Ajeet Kumar
Advanced Photonics Simulator Lab,
Department of Applied Physics,
Delhi Technological University
Delhi-110042, India
ajeetdph@dtu.ac.in

Dinesh Kumar Vishwakarma
Biometric Research Laboratory, Department of Information Technology,
Delhi Technological University
Delhi-110042, India
dvishwakarma@gmail.com

2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT) | 979-8-3503-3936-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/APSIT58554.2023.10201711

***Abstract*—Multimodal aspect-based sentiment recognition (MABSR) is a recently developed task in sentiment recognition that tries to assess the sentiment associated with text and image pairings by generally extracting the polarity terms from the pairs. Both the pipeline and the unified transformer based technique, which employs the cross-attention only mechanism, have been widely utilized in recent works. However, the alignment between text and picture is not openly and reliably included in these approaches. There is still a minimum threshold of aligned image-text pairings needed for supervised fine-tuning of said universal transformers for MABSR. Motivated by this observation and inspired by the various attention-only mechanisms, we analyze MABSR and propose an attention-free encoder-based transformer architecture. Dual attention-free based backbone encoder models with cross-modal symmetry are utilized in this work. To improve cross-modal performance, we include two new subtasks: aspect-only extraction and polarity feature representation alignment. This motivates both encoders to provide more precise depictions of multiple modalities.**

Keywords—Multimodal Aspect Based Sentiment Recognition (MABSR), Multi-task learning, Attention-free encoders, Cross-Modal Fusion.

I. INTRODUCTION

Multimodal aspect-based sentiment recognition (MABSR) refers to the analysis of sentiments along with the extraction of the aspect terms within multiple modalities. Our immersive awareness allows us to see the environment in many different ways, including via sight, sound, touch, and taste. In most cases, one aspect of a multimodal experience will correlate with another. Images, for instance, typically have tags and explanatory text attached to them, and vice versa to better communicate the speaker's primary objective. Communication via social media posts has become increasingly multimodal i.e., via text and visual content due to the ubiquitous accessibility of cell devices with digital cameras. One must be able to understand these multimodal attributes jointly to effectively utilize the existing approaches.

Fig 1 (a) shows an example: *That boy Fellaini, what a talent in MMA; maybe dirty elbowing shithouse of a footballer though!* An ideal system is anticipated to yield three aspect-sentiment pairs from this sentence and image, i.e., (*Fellaini*, positive), (*MMA*, neutral) and (*footballer*, negative). Firstly, if the input is using text modality only at the inference step, the model tends to predict (*Fellaini*, neutral) and (*MMA,* positive*)*. When the visual modality is also included, the context of the text tends to be neutral.

Fig 1. Two examples of Aspect Based Multi-Modal Sentiment Recognition.

Fig 1 (b) shows another example: *Former top John McCain aide says he'd back Hillary Clinton over Donald Trump*. It is difficult to infer from the textual modality only about the polarity sentiment of the three personas in context of the sentence.

Many former studies have tried a variety of ways to merge the acquired representations of both modalities as the final classification features. To establish a text-image fusion and generate vision-aware-context and vice versa, researchers have utilized various techniques, however the image-text pairings convey distinct types of information as there is not an equal proportion of knowledge in each modality. With the gradients from the text modality tending to dominate over a weaker modality in terms of information, it will be challenging to train the model. Estimating the attention distribution superficially makes encoding explicit cross-modal information challenging.

Key Contributions:

- For the task of MABSR, we present a dual-encoder technique that incorporates cross-modal information. Instead of extracting ROI features, the MLP-Mixer [1] architecture is utilized, which tokenizes the visual input by dividing it into a series of patches. The textual and visual modalities were encoded using FNet [2] and, MLP-Mixer [1] which are both initialized from pre-trained checkpoints. We used a multitask learning framework comprising a few subtasks and these are — aspect-only extraction, polarity feature integration and co-attention fusion to fuse the acquired information.

979-8-3503-3936-9/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:42:43 UTC from IEEE Xplore. Restrictions apply.

- To evaluate the efficacy of proposed framework, we utilised two comparable data sets, Twitter-2015 and Twitter-2017, which mainly contain social media opinions. The empirical results demonstrate that the proposed framework for MABSR task surpasses both the current unimodal and multimodal approaches at lower compute and more efficient memory requirements. Further evaluation of the effects on various subtasks revealed that each subtask plays an indispensable role in performance enhancement.

The remaining work is organised as follows: Section II presents former related studies on the MABSR task, Section III describes the methodology employed and further implementation details. Section IV discussed the experimental results, and the conclusion is summarised in Section V.

II. Related Work

A. Unimodal Aspect-Based Sentiment Recognition

1) Based on Textual Content

Aspect-based sentiment recognition (ABSR) is a branch of sentiment analysis that aims to identify polarity towards specific aspects in a given context. [3]; [4]; [7] were among the early works which explored ABSR. They proposed various techniques such as utilizing sentiment lexicons, syntactic patterns, and word embeddings for target-dependent sentiment classification via recurrent neural networks due to its ability to extract semantic features from texts.

Recent studies have also focused on deep learning models that utilize attention mechanisms to dynamically assign weights to relevant aspects and corresponding sentiment expressions. [6] proposed an LSTM based approach; [7] introduced an 'Interactive Attention Network (IAN)' that incorporated attention mechanisms. Memory networks, have also been explored for fine-grained opinion mining. [8] proposed a 'Target-Sensitive Memory Network (TSMN)' that employs a hierarchical memory mechanism for targeted storage and retrieval of aspect-specific sentiment information. [9] proposed a 'Latent Opinion model' that integrates a latent opinion variable into the sentiment recognition framework. [10] discussed a GCM-IAN model that incorporated granular computing to enhance the interpretability of parameters.

Additonally, transformer-based architectures have also shown superior performance in various NLP tasks, [11] proposed a BERT-based approach that utilizes a self-attention mechanism that integrates field-specific expertise through refinement of the pre-trained model. Although the above proposed models have shown promise, challenges in sentiment recognition still persist i.e. detecting sarcasm and irony in abundance of shifting trends to image supported text simply is a challenging task for current models due to the need for a nuanced understanding of context and social cues.

2) Based on Visual Content

Visual sentiment recognition (VSR) is also a research area that belongs to a broader range of sentiment recognition tasks. The objective is to comprehend human viewpoints, and stances via visual media, such as pictures and videos. A prevalent method for VSR entails examining facial expressions. [12]; [13]; [14] suggested a deep network in which domain features are transferred to perform well on smaller datasets with greater accuracy. [15] utilised skip-gram and autoencoders for enhanced VSR feature learning. Another approach would be to focus on localized features, [16]; [17] proposed an attention mechanism that captures fine-grained information by attending to local image regions. [18] proposed predicting visual sentiments through the automatic discovery of affective regions. Additionally, [19] proposed a 'context-dependent sentiment recognition method' which employed text and visual elements to convey a videos' sentiment. As the field continues to develop, one of the major challenges hindering progress is the absence of large-scale accurately labelled datasets. The other is the need for more robust and interpretable models that can capture complex relationships.

B. Multimodal Aspect Based Sentiment Recognition

Aspect-based multimodal sentiment recognition seeks to classify the sentiment conveyed in a text or image by examining its various aspects [20]; [21] conducted early experiments in the field, the latter proposed 'Convolutional Multiple Kernel Learning (MKL)' approach that aims to determine the optimal kernel for each modality. [22] discussed 'word-level fusion and reinforcement learning-based' approach for sentiment recognition. [23] proposed MultiSentiNet, a deep semantic network that utilized a combination of LSTM and CNN networks for feature extraction.

Further studies for MABSR were conducted by [24]; [25] as they studied a 'multi-interactive memory network' to encode text and image information. [26] utilised a gated attention mechanism; [27] proposed utilizing 'attention capsule extraction' via a fusion network with entity-level classification for the text and their corresponding images. Researchers have also investigated the utilization of pre-trained large language [11] and vision models for MABSR as [28] introduced VL-BERT and [29] proposed UNITER i.e. both architectures simultaneously learns visual and textual representations and can be fine-tuned for MABSR task.

Existing approaches typically consist of two pipeline subtasks: 1) multimodal aspect term retrieval; 2) sentiment classification. The former aims at recognizing all aspect terms from sequences, whereas the latter attempts to categorize each identified aspect term with its associated sentiment. However, the pipeline approach disregards the inherent connection between the two tasks and is susceptible to information loss. Mixing diverse modalities to boost performance is logical, but overcoming the noise and conflicts across such modalities proves difficult.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:42:43 UTC from IEEE Xplore. Restrictions apply.

III. Proposed Methodology

Fig. 2 shows the overview of our proposed methodology. We define the above problem as a token classification task in terms of polarity and corresponding aspects. In the following subsections, we first introduce our separate input representations and their encoders as feature extractors for both modalities, followed by the details of our cross-modal feature alignment and subsequent subtasks involved.

A. Image Representation

For visual modality, unlike many existing transformer [30] derivatives, we employ MLP-Mixer to extract spatial features from an input image. The input image is first sliced into a sequence of patches $s = [\, s_1, s_2, \ldots\ldots, s_I]$ $where\ s \in R^{I\times(P^2\times C)}$, P^2 is the resolution per patch, C is the number of color channels, and $I = \mathbb{H} \times \mathbb{W} \,/\, P^2$ is the resulting number of non-overlapping patches [1]. Each patch is then flattened and prefixed with a special token, i.e., [*CLS*], followed by a simple linear projection $S \in R^{(P^2\times C)\times d_h}$. The resulting patch embeddings $S_{emb} \in R^{I\times d_h}$ can be formulated as, $S_{emb} = [\, s_1 S, s_2 S, \ldots\ldots, s_I S]$ where d_h is the dimensionality. Our vision encoder does not need position embeddings.

B. Text Representation

For textual modality, we first tokenize the input text by the Unigram Tokenizer [31] utilized in the FNet model to obtain a sequence of tokens embeddings $\bar{t} \in R^{(L+1)\times d_h}$ with a word embedding matrix $T \in R^{L\times|\bar{g}|}$ as follows, $\bar{T} = \left[\, t_{[CLS]}, t_1 T, t_2 T, \ldots\ldots, t_L T, t_{[SEP]}\right] + T^{pos}$ where $T^{pos} \in R^{(L+1)\times d_h}$, $|\bar{g}|$ represent the position embeddings, and the number of the vocabulary items respectively. Here, the $[CLS]$ and $[SEP]$ tokens correspond to the classification and end-of-sequence tokens in the text encoder. No supplementary embeddings were applied to the input text modality, as it did not yield any improvement to the empirical results.

C. Encoders

Both Fnet and MLP-Mixer are constructed from a series of layers of stacked transformer-like mechanism blocks, which provided behaviour of mixing tokens without requiring the attention mechanism. Fnet replaces the general self-attention sublayer [30] within former Transformer architecture with a Fourier sublayer, which applies a 2-D discrete fourier transform (DFT) (from Eq. 1) to its input token embeddings across sequences of length F_{seq} and hidden dimension, d_h.

$$X_K = \sum_{l=0}^{L-1} x_l e^{-(\frac{2\pi i}{L})lk}, 0 \le k \le L-1 \quad (1)$$

$$y = \Re(F_{seq}(F_{d_h}(x))) \quad (2)$$

Here $\Re$ is the real part of the extracted composite output.

$$\tilde{z}^{(n)} = LayerNorm\left(DFT\ Mixer\ (Z^{n-1})\right) + Z^{n-1} \quad (3)$$

$$Z^n = LayerNorm\left(MLP\ (\tilde{z}^{(n)})\right) + Z^{n-1} \quad (4)$$

Similarly, MLP-Mixer replaces the self-attention block responsible for mixing features between different spatial locations with two identical sized MLP blocks.

The former is responsible for token-mixing and the latter is responsible for channel-mixing. The final MLP layer in either encoders comprises of two dense layers with a GELU non-linear activation. Before both the token mixing modules and final MLP block, layer normalization (LayerNorm) was applied respectively for both FNet or MLP-Mixer mechanism (from Eq 4; Eq. 6). The token mixer block represents either Fast Fourier Transform approach from FNet or all fully connected MLP layer approach from MLP-Mixer (from Eq. 3; Eq.5).

$$\tilde{z}^n = TokenMixer\left(LayerNorm\ (z^{(n-1)})\right) + z^{(n-1)} \quad (5)$$

$$z^n = ChannelMixer\left(LayerNorm\ (\tilde{z}^{(n)})\right) + \tilde{z}^{(n)} \quad (6)$$

The final output of our encoders are hidden representations $Z_s^{(N)} = [\hat{s}_1, \hat{s}_2, \ldots\ldots, \hat{s}_I]$ and $Z_T^N = [\hat{t}_1, \hat{t}_2, \ldots\ldots, \hat{t}_L]$ (Eq. 7a; 7b) at the L-th layer, which is last in our case. It is used as input for multi-task learning and the final extraction. For all the experiments, the parameters of FNet and MLP-Mixer were respectively initialized from pretrained *fnet-base* and *mixer-b16-224-in21k* checkpoints. The hidden size d_h is 768, the number of encoders layers, N is 12, Patch Size, P is 14, and final MLP layer hidden dimension is 3072.

D. Cross-Modal Fusion:

With the goal of aligning the characteristics of both the visual and language modalities, we propose a cross-modal fusion layer to train both the image and language encoders for the final extraction. Our approach involves two subtasks: aspect-only extraction and co-attention exchange with polarity feature alignment.

1) Text extraction:

The textual representation received from text encoder i.e.

$$Z_T^N = \left[\hat{t}_{[CLS]}, \hat{t}_1, \hat{t}_2, \ldots\ldots, \hat{t}_L, \hat{t}_{[SEP]}\right] \quad (8)$$

was input to a dense layer with softmax activation to estimate auxiliary tags for the tokens, defined as,

$$\hat{y}_n = Softmax(w^t \hat{t}_n + b^t) \quad (9)$$

where $w^t \in \mathbb{R}^{K\times dh}$ and $b^t \in \mathbb{R}^K$ are trainable parameters, and K are the number of the prospective tags for either subtask. Given a training dataset of {x(j), y(j)}, the loss function used is Dice Loss proposed in [32],

$$L^{DL} = \frac{1}{J\times N} \sum_j^{j\in\{0,1,2\}} \sum_i \left[1 - \frac{2p_{ij}\, y_{ij} + \gamma}{p_{ij}^2 + y_{ij}^2 + \gamma}\right] \quad (10)$$

The BIO schema was utilised for token classification. We employed 3 tags, *I-ASP, O, B-ASP* for subtask 1 and a reduced set of 5 tags, specifically *I, O, B-POS, B-NEU, and B-NEG*, as opposed to the previous usage of 7 tags for subtask 2 and 3. For instance, the pattern of {*B-NEG, I-NEG*} can be compressed to {*B-NEG, I*}, trimming the number of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:42:43 UTC from IEEE Xplore. Restrictions apply.

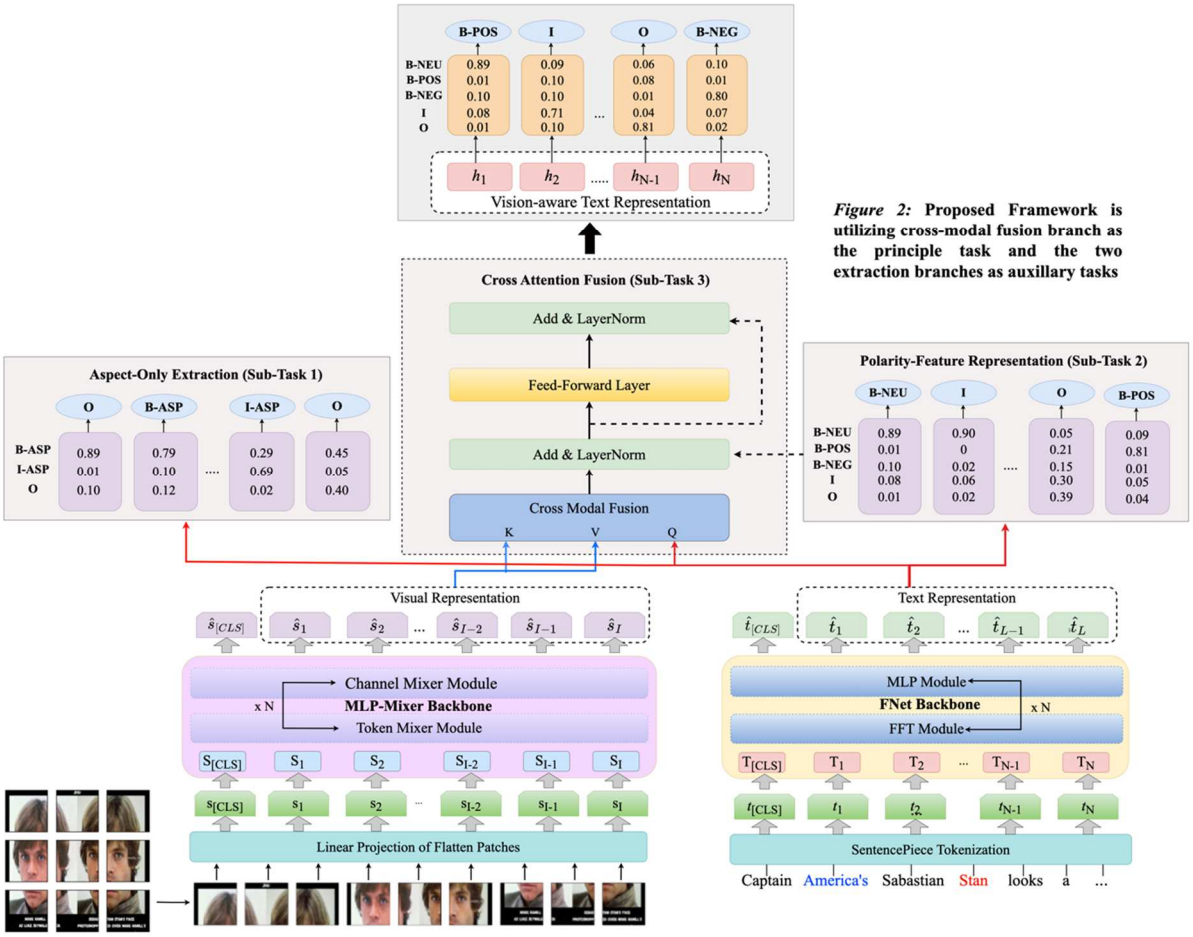


Figure 2: **Proposed Framework is utilizing cross-modal fusion branch as the principle task and the two extraction branches as auxillary tasks**

Fig. 2. Proposed Framework

classes K. This can reduce the overall forecasting error resulting from the classes K in the task.

E. Vision-aware Text Extraction:

After further investigation, multi-headed cross-attention was applied to integrate the textual and visual features where the text representation (from eq. 8) is regarded as the query, while the image representation (from eq. 7a) was utilised as the key and the value,

$$Atte_u\left(\mathcal{Z}_T^{(N)}, \mathcal{Z}_S^{(N)}, \mathcal{Z}_S^{(N)}\right) = softmax\left(\frac{\left(W_Q^U \mathcal{Z}_T^{(N)}\right)\left(W_K^U \mathcal{Z}_T^{(N)}\right)}{\sqrt{d_h/_u}}\right)\left(\left(W_V^U \mathcal{Z}_S^{(N)}\right)\right) \quad (11)$$

where $W_Q^U \in \mathbb{R}^{d_h/u \times L}$ and $\{W_K^U \quad W_V^U\} \in \mathbb{R}^{d_h/u \times I}$ are matrices of the query, key and value [30]. With multi-headed cross-fusion, the final representation of vision-aware text extraction $\bar{p} = [p_1, p_2, \ldots, p_n]$ can be formulated as,

$$\bar{p} = W^p[Atte_1, Atte_2, \ldots, Atte_u]^T \quad (12)$$

where $W^p \in R^{d_h \times d_h}$ refer to the parameter matrix for the multi-headed cross fusion. By passing a fully connected dense layer and two-layer normalization with two skip connections, the resulting representation is formulated as $\hat{p} = [\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_L]$. The skip connections incorporated the text-based representation to maintain a compatible size of representation. Different from the aspect or polarity-only extraction subtasks, the output layer uses CRF Score [33] estimation as input to our loss function which is the negative log-probability of the ground truth label,

$$L^{CRF} = -\frac{1}{J}\sum_{j=1}^{J}\left(s\left(x^{(j)}, y^{(j)}\right) - logadd_{y' \in Y_x^{(j)}} e^{s\left(x^j, y'^{(j)}\right)}\right)$$

where L^{CRF} $(Eq.\,13)$ is the CRF loss and $s\left(x^{(j)}, y^{(j)}\right)$ is the estimated CRF score. The main purpose is to integrate the three subtasks into a cohesive whole, as outlined below,

$$L = L^{CRF} + \theta L_1^{DL} + \delta L_2^{DL} \quad (14)$$

where θ and δ are hyper-parameters that regulate the impact of individual subtasks. The output of solely vision-aware context extraction ($\theta = 0$, $\delta = 1$) was utilised for inference.

IV. Experimental Results:

A. Datasets and its brief description and statistics:

To evaluate the performance of our proposed attention-free dual encoders with cross-modal fusion, we utilized two MABSR benchmark datasets, mainly consisting of reviews on Twitter. These datasets are Twitter-2015 and Twitter-2017 provided by [34]; [35]. These were annotated with the sentiment for each aspect and can be accessed from —

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:42:43 UTC from IEEE Xplore. Restrictions apply.

https://github.com/jefferyYu/TomBERT. An aspect is deemed correctly identified only when the aspect term and its polarity match the ground-truth aspect term as well as polarity. We evaluated our approach on the precision, recall, and micro F1-score metrics.

Table 1: Statistical characteristics of respective dataset.

Datasets	Positive Aspects	Negative Aspects	Neutral Aspects	Total Aspects
Twitter-15	1548	630	3169	5347
Twitter-17	2516	728	2728	5972

B. Implementation details

The Hyperparameters were fine-tuned through grid-search based on validation set performance in all experiments. The hidden size d_h is 768 for the visual and language encoders. The cross-modal fusion module has 8 heads. The AdamW optimizer was utilised to update all trainable parameters with a primary learning rate of $1e^{-4}$ and warmup of 0.1. The max length and batch size were set to 80 and 64 limited by the GPU (Nvidia P100 with 16 GB VRAM), respectively. For training, we leveraged an early stopping strategy with a patience of 15 to 25 epochs to avoid overfitting. Both θ and δ belonged to the prospective set of {0.2, 0.8, 0.3, 0.7, 1.0}.

C. Results

Table 2 and **Table 3** llustrate the experimental results for the task of MABSC. **Fig. 3** gives the comparision of our proposed framework with the previous state of the art approaches for both the datasets. The empirical results for both Twitter-15 and Twitter-17 datasets including comparative analysis with [36]; [25]; [27]; [37]; [38].

Table 2: Results analysis on Twitter-15 dataset with our proposed framework.

Modalities	Approaches	Twitter-15		
		P_r	R_C	F_1
Text only	TomBERT	-	-	72.6
	ESAFN	-	-	-
	EF-Net	-	-	67.30
	Proposed (Ours)	**74.20**	**73.26**	**73.53**
Text + Image	ESAFN	-	-	67.37
	TomBERT	-	-	71.27
	EF-Net	-	-	67.90
	ModalNet	-	-	72.50
	HIMT	-	-	73.68
	CapBERT	-	-	73.09
	Proposed (Ours)	**73.36**	**72.54**	**72.97**

Table 3: Results analysis on Twitter-17 dataset with our proposed framework.

Modalities	Approaches	Twitter-17		
		P_r	R_C	F_1
Text	TomBERT	-	-	67.48
	ESAFN	-	-	-
	EF-Net	-	-	62.21
	Proposed (Ours)	**74.20**	**73.26**	**73.53**
Text + Image	TomBERT	-	-	68.04
	ESAFN	-	-	64.22
	EF-Net	-	-	65.32
	ModalNet	-	-	69.19
	HIMT	-	-	69.16
	CapBERT	-	-	70.20
	Proposed (Ours)	**72.12**	**71.40**	**71.76**

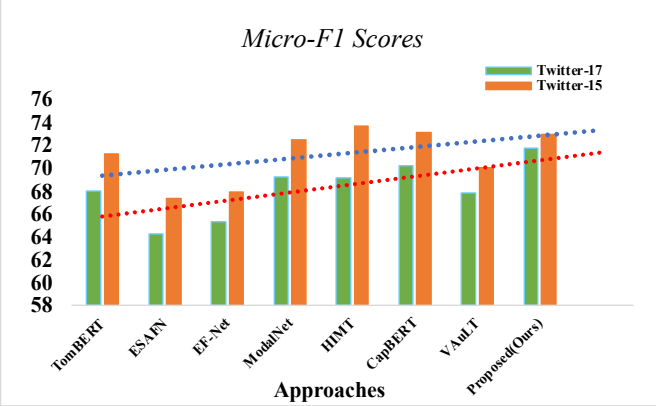


Fig. 3. Comparitive graphical analysis with previous cutting-edge methods using F1 score metric.

V. Conclusion

This work proposes a dual-encoder mechanism with cross-modal fusion for encoding text-image features into the representations for MABSR tasks. A multitask learning architecture containing three subtasks was applied to coalesce both text and image modalities. Empirical experiments demonstrate the effectiveness of the proposed framework improved the performance for MABSR tasks verified via Twitter-2015 and Twitter-2017 datasets. In addition, future work can extend the proposed framework to more multi-modal tasks, such as multi-modal machine reading comprehension and affective computing.

References

[1] I. O. Tolstikhin *et al.*, "MLP-Mixer: An all-MLP Architecture for Vision," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, pp. 24261–24272. [Online]. Available: https://proceedings.neurips.cc//paper/2021/hash/cba0a4ee5ccd02fda0fe3f9a3e7b89fe-Abstract.html

[2] J. Lee-Thorp, J. Ainslie, I. Eckstein, and S. Ontanon, "FNet: Mixing Tokens with Fourier Transforms," in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 4296–4313. doi: 10.18653/v1/2022.naacl-main.319.

[3] J. Wagner *et al.*, "DCU: Aspect-based Polarity Classification for SemEval Task 4," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland: Association for Computational Linguistics, Aug. 2014, pp. 223–229. doi: 10.3115/v1/S14-2036.

[4] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews," in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:42:43 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland: Association for Computational Linguistics, Aug. 2014, pp. 437–442. doi: 10.3115/v1/S14-2076.

[5] D.-T. Vo and Y. Zhang, "Target-dependent twitter sentiment classification with rich automatic features," in *Proceedings of the 24th International Conference on Artificial Intelligence*, in IJCAI'15. AAAI Press, Jul. 2015, pp. 1347–1353.

[6] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for Aspect-level Sentiment Classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 606–615. doi: 10.18653/v1/D16-1058.

[7] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive Attention Networks for Aspect-Level Sentiment Classification," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, Aug. 2017, pp. 4068–4074. doi: 10.24963/ijcai.2017/568.

[8] S. Wang, S. Mazumder, B. Liu, M. Zhou, and Y. Chang, "Target-Sensitive Memory Networks for Aspect Sentiment Classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 957–967. doi: 10.18653/v1/P18-1088.

[9] B. Wang and W. Lu, "Learning Latent Opinions for Aspect-level Sentiment Classification," *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, Art. no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.12020.

[10] H. Li, T. Yuan, H. Wu, Y. Xue, and X. Hu, "Granular computing-based multi-level interactive attention networks for targeted sentiment analysis," *Granul. Comput.*, vol. 5, no. 3, pp. 387–395, Jul. 2020, doi: 10.1007/s41066-019-00163-9.

[11] C. Sun, L. Huang, and X. Qiu, "Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 380–385. doi: 10.18653/v1/N19-1035.

[12] C. Xu, S. Cetintas, K.-C. Lee, and L.-J. Li, "Visual Sentiment Prediction with Deep Convolutional Neural Networks." Nov. 20, 2014. doi: 10.48550/arXiv.1411.5731.

[13] Q. You, J. Luo, H. Jin, and J. Yang, "Joint Visual-Textual Sentiment Analysis with Deep Neural Networks," in *Proceedings of the 23rd ACM international conference on Multimedia*, in MM '15. Association for Computing Machinery, Oct. 2015, pp. 1071–1074. doi: 10.1145/2733373.2806284.

[14] Q. You, J. Luo, H. Jin, and J. Yang, "Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, in AAAI'15. Austin, Texas: AAAI Press, Jan. 2015, pp. 381–388. [Online]. Available: http://arxiv.org/abs/1509.06041

[15] C. Baecchi, T. Uricchio, M. Bertini, and A. Del Bimbo, "A multimodal feature learning approach for sentiment analysis of social network multimedia," *Multimed. Tools Appl.*, vol. 75, no. 5, pp. 2507–2525, Mar. 2016, doi: 10.1007/s11042-015-2646-x.

[16] Q. You, L. Cao, H. Jin, and J. Luo, "Robust Visual-Textual Sentiment Analysis: When Attention meets Tree-structured Recursive Neural Networks," in *Proceedings of the 24th ACM international conference on Multimedia*, in MM '16. Association for Computing Machinery, Oct. 2016, pp. 1008–1017. doi: 10.1145/2964284.2964288.

[17] Q. You, H. Jin, and J. Luo, "Visual Sentiment Analysis by Attending on Local Image Regions," *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, Art. no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.10501.

[18] J. Yang, D. She, M. Sun, M.-M. Cheng, P. L. Rosin, and L. Wang, "Visual Sentiment Prediction Based on Automatic Discovery of Affective Regions," *IEEE Trans. Multimed.*, vol. 20, no. 9, pp. 2513–2525, Feb. 2018, doi: 10.1109/TMM.2018.2803520.

[19] S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, and L.-P. Morency, "Context-Dependent Sentiment Analysis in User-Generated Videos," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 873–883. doi: 10.18653/v1/P17-1081.

[20] S. Poria, E. Cambria, and A. Gelbukh, "Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-level Multimodal Sentiment Analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 2539–2544. doi: 10.18653/v1/D15-1303.

[21] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional MKL Based Multimodal Emotion Recognition and Sentiment Analysis," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, Spain: IEEE, Dec. 2016, pp. 439–448. doi: 10.1109/ICDM.2016.0055.

[22] M. Chen, S. Wang, P. P. Liang, T. Baltrušaitis, A. Zadeh, and L.-P. Morency, "Multimodal sentiment analysis with word-level fusion and reinforcement learning," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, in ICMI '17. New York, NY, USA: Association for Computing Machinery, Nov. 2017, pp. 163–171. doi: 10.1145/3136755.3136801.

[23] N. Xu and W. Mao, "MultiSentiNet: A Deep Semantic Network for Multimodal Sentiment Analysis," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, in CIKM '17. Singapore: Association for Computing Machinery, Nov. 2017, pp. 2399–2402. doi: 10.1145/3132847.3133142.

[24] N. Xu, W. Mao, and G. Chen, "Multi-Interactive Memory Network for Aspect Based Multimodal Sentiment Analysis," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, Art. no. 01, Jul. 2019, doi: 10.1609/aaai.v33i01.3301371.

[25] J. Yu, J. Jiang, and R. Xia, "Entity-Sensitive Attention and Fusion Network for Entity-Level Multimodal Sentiment Classification," *IEEEACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 429–439, Dec. 2019, doi: 10.1109/TASLP.2019.2957872.

[26] A. Kumar and J. Vepa, "Gated Mechanism for Attention Based Multi Modal Sentiment Analysis," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020, pp. 4477–4481. doi: 10.1109/ICASSP40776.2020.9053012.

[27] D. Gu *et al.*, "Targeted Aspect-Based Multimodal Sentiment Analysis: An Attention Capsule Extraction and Multi-Head Fusion Network," *IEEE Access*, vol. 9, pp. 157329–157336, Nov. 2021, doi: 10.1109/ACCESS.2021.3126782.

[28] W. Su *et al.*, "VL-BERT: PRE-TRAINING OF GENERIC VISUAL-LINGUISTIC REPRESENTATIONS," in ICLR. 2020.

[29] Y.-C. Chen *et al.*, "UNITER: UNiversal Image-TExt Representation Learning," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., in Lecture Notes in Computer Science, vol. 12375. Springer International Publishing, 2020, pp. 104–120. doi: 10.1007/978-3-030-58577-8_7.

[30] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, Long Beach, CA, USA: Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[31] T. Kudo, "Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 66–75. doi: 10.18653/v1/P18-1007.

[32] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice Loss for Data-imbalanced NLP Tasks," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online: Association for Computational Linguistics, Jul. 2020, pp. 465–476. doi: 10.18653/v1/2020.acl-main.45.

[33] R. Panchendrarajan and A. Amaresan, "Bidirectional LSTM-CRF for Named Entity Recognition," in *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong: Association for Computational Linguistics, 2018. [Online]. Available: https://aclanthology.org/Y18-1061

[34] Q. Zhang, J. Fu, X. Liu, and X. Huang, "Adaptive Co-attention Network for Named Entity Recognition in Tweets," *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.11962.

[35] D. Lu, L. Neves, V. Carvalho, N. Zhang, and H. Ji, "Visual Attention Model for Name Tagging in Multimodal Social Media," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 1990–1999. doi: 10.18653/v1/P18-1185.

[36] J. Yu and J. Jiang, "Adapting BERT for Target-Oriented Multimodal Sentiment Classification," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, Macao, China: International Joint Conferences on Artificial Intelligence Organization, Aug. 2019, pp. 5408–5414. doi: 10.24963/ijcai.2019/751.

[37] Z. Zhang, Z. Wang, X. Li, N. Liu, B. Guo, and Z. Yu, "ModalNet: an aspect-level sentiment classification model by exploring multimodal data with fusion discriminant attentional network," *World Wide Web*, vol. 24, no. 6, pp. 1957–1974, Nov. 2021, doi: 10.1007/s11280-021-00955-7.

[38] J. Yu, K. Chen, and R. Xia, "Hierarchical Interactive Multimodal Transformer for Aspect-Based Multimodal Sentiment Analysis," *IEEE Trans. Affect. Comput.*, pp. 1–1, 2022, doi: 10.1109/TAFFC.2022.3171091.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:42:43 UTC from IEEE Xplore. Restrictions apply.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Blockchain-Enabled Healthcare Records Management: A Survey of Implementation Strategies

1st Divya Kuntal
Department of Information Technology
Delhi Technological University
Delhi, India
divyakunttall@gmail.com

2nd Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological University
Delhi, India
dinesh@dtu.ac.in

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205912

***Abstract*—Electronic Health Records (EHR) is a technology that enables users to monitor their medical data. Over several healthcare organizations, it maintains computerized records. Enterprise-wide data exchanges and other networking exchanges and technologies are used to exchange records. People today demand fast access to their medical records. The privacy and security of patient medical records are a concern because the health industry has rapid access to data. Thus, a blockchain-based solution can help to solve this problem. A blockchain network is used in healthcare to store and distribute health information amongst hospitals, diagnostic laboratories, pharmaceutical firms, and physicians. This study uses Blockchain Technology (BT) to analyze electronic healthcare data management. Insightful analysis of patient data is made possible by such a technology, which medical institutions can utilize. After an analysis of the survey of the relevant literature, a comparison of the many earlier studies is presented. Encryption, the keyword (KW) search, server and user verifiability, storage integrity, and dynamic updates are all aspects that have been compared. After examining. the author Tan et al., (2022) agrees with all the key features such as server-side verifiability, encryption, KW search, user-side verifiability, storage integrity, and dynamic updating**

***Keywords*— Blockchain, EHR, Medical records, Personal Health Record (PHR)**

I. INTRODUCTION

Healthcare providers, administrators, and researchers increasingly rely on Electronic Health Records (EHRs), which are computerized compilations of healthcare operations and assessments. [1]. Structured and unstructured data are both present in electronic health records [2]. Diagnoses, medications, and test values are all examples of the types of structured data found in EHRs. However, examples of unstructured data include clinical documentation such as notes, and discharge summaries created by healthcare professionals. When doctors enter their notes on a patient's condition into an electronic health record (EHR), they do it in free text. Electronic health record (EHR) adoption has skyrocketed in recent years. It has risen drastically in the US, going from 10% to almost 96% in just ten years. The sum is up by well over 85 percent [3]. The same pattern holds true for Australia's GP offices, major hospitals, and medical service providers [4,5]. The widespread use of EHRs has resulted in a data volume large enough to be classified as "Big Data," which includes the processing and use of the vast amounts of information stored in EHRs. There is a need to invent technological devices that can organize, evaluate, and recognize patterns within this information because the ability of the human intellect to analyze, understand, and interpret information is limited. The next phase in creating an EHR ecosystem is to conduct an analysis of data on EHRs big data, which includes the implementation of information mining and processing of natural language techniques. To deal with the explosion of free-form text in the medical field, researchers have turned increasingly to cutting-edge Machine Learning (ML) methods [6,7]. Many medical and healthcare contexts [8] include the diagnosis of heart disease and other cardiovascular issues [9]. Blockchain is an exciting new technology with the potential to revolutionize healthcare data administration by enabling greater data efficiency and ensuring trust [10-15]. Decentralized storage, transparency, data integrity, verification, data access flexibility, connectivity, and security are remarkable and built-in properties that allow BT to employ extensively in the healthcare industry [16-17]. The healthcare systems possible by BT are shown in Figure 1 [18-19].

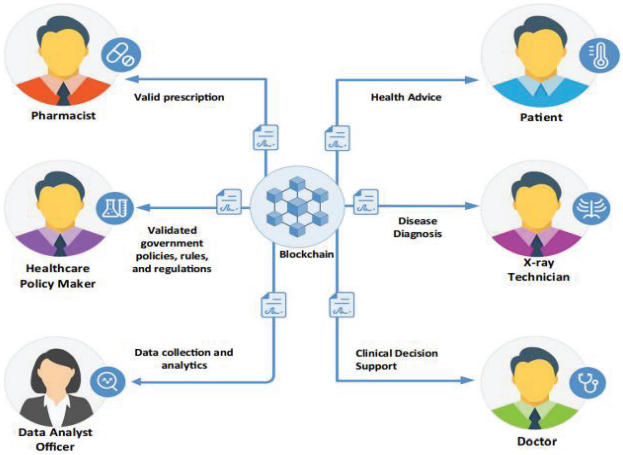


Fig. 1. Blockchain-enabled healthcare systems [20].

Blockchain data management systems are instead addressing the integration of data silos. Despite several recognized eHealth interoperability strategies, health record data is still compartmentalized inside hospitals [21]. Interoperability worldwide is not yet a reality, even if the size of these silos expands (to include the whole European continent or the United States, for example) [22]. In such a case, the use of BT can aid in establishing sharing practices to solve the obstacles of data dissemination [23]. New possibilities for everyday health data management and for patients to easily access and exchange health data are

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:29 UTC from IEEE Xplore. Restrictions apply.

opening up due to developments in EHR collecting, patient data privacy protection policies, and cloud healthcare data storage (Table 1).

TABLE I. BLOCKCHAIN-BASED HEALTH RECORDS MANAGEMENT FIRMS [24]

	Blockchain Company	
	Name	Country
EMR data management	PokitDoc	USA
	Gem	USA
EHR data management	Medicalchain	USA
	Health Wizz	USA
	Curisium	USA
PHR data management	Hearthy	Spain
	Medcredits	USA
	MyClinic	UK

Note: EMR: Electronic Medical Record, PHR: Personal Health Record.

The rest of this work is structured as follows. The authors provide a comparison table and a literature review of surveys in Section 2. The authors offer a comparative literature review analysis with the graph in Section 3. Finally, they stated the conclusions in Section 4.

A. Blockchain-based Management Applications

Opportunities for better health data management and easier patient access to and sharing of health data have arisen with developments in EHR, patient data privacy protection policies, and cloud healthcare data storage [25].

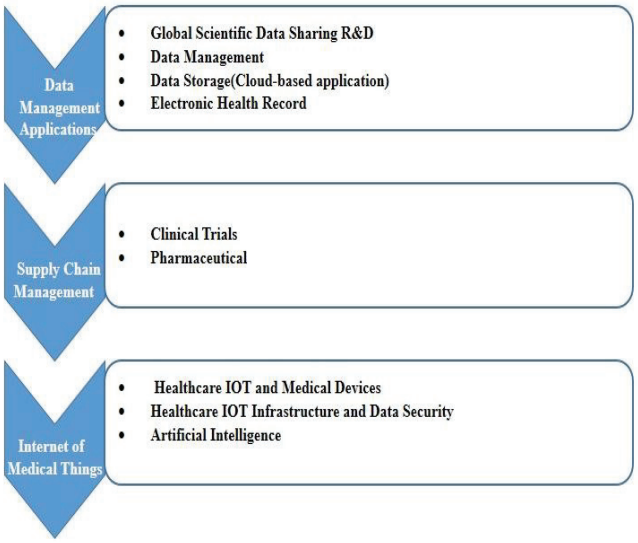


Fig. 2. Healthcare blockchain applications [26].

Applications of BT in healthcare are shown in Figure 2. BT can tackle these crucial challenges robustly and efficiently, which is particularly important in the healthcare industry, which relies heavily on secure data storage and transactions and the management of their seamless integration.

TABLE II. THE POTENTIAL OF BT IN THE MEDICAL INDUSTRY [27].

Applications	Summary
EHR	The digital EHR stored on the distributed ledger of a permitted blockchain maintains its integrity without the need for any intermediaries from the moment data is generated until it is retrieved.
Clinical Research	Blockchain provides a trustless, immutable structure for sharing data in clinical trials. It allows investigators to communicate data with one another in a safe environment.
Medical Fraud Detection	As a distributed ledger that cannot be altered once recorded, blockchain facilitates fraud detection and prevents unauthorized duplicates or alteration of transactions.
Pharmaceutical Industry and Research	Blockchain's sophisticated tracking capabilities are used in the pharmaceutical industry to maintain tabs on the whole supply chain, ensuring that pharmaceutical counterfeiting and theft are prevented at every level.

The seven stages of the blockchain-based healthcare data management process are shown in Figure 3. Data management, data sharing, data storage (such as cloud-based apps), and EHR are all examples of applications built on BT that will be covered in further depth below. Data Management, data sharing, data storage (such as cloud-based apps), and EHR are all examples of applications built on BT that will be covered in further depth below.

1) Data Management

The fact that many businesses, particularly healthcare institutions, are data-driven and that the amount of data created now and in the future, thanks to technologies like the IoT, is rising exponentially, there are persistent breaches of data security and privacy [28]. Therefore, many institutions' credibility and financial resources have been severely damaged. Health data users should be separated into categories based on their functions, with access controlled according to the level of authorization granted to each category. BT can be employed to guarantee such access is secure and discreet.

2) Electronic Health Record

Paper-based medical records are cumbersome because they need accurate tracking of a patient's health condition over time [29]. Also, they are vulnerable to inaccurate information, which can lead to inappropriate treatment of patients. The development of Information Technology (IT) allowed for the implementation of EHRs, significantly reducing the need for such initiatives. Access to EHR enables medical practices to increase treatment quality [30]. Additionally, EHR allows for improved disease management and higher levels of preventative treatment. The digital

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:29 UTC from IEEE Xplore. Restrictions apply.

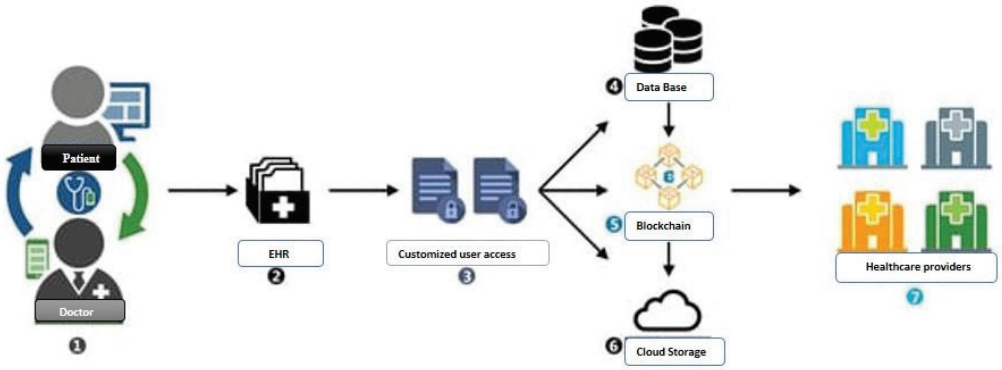


Fig. 3. Blockchain-based healthcare data management [26].

record also enables improved decision-support features and enhanced teamwork among medical professionals. Consequently, its importance is becoming more widely acknowledged in the medical field [31].

II. LITERATURE REVIEW

In this section, the author studies about the previous research work based on electronic healthcare records (EHR) management using blockchain technologies (BT).

Tan et al., (2022) [32] investigate using BT with smart contract implementations to build a decentralized, verifiable, and unchangeable database. The suggested approach allows for immutable storage, real-time changes to EHR, and the protection of patient privacy when using an outsourced service. The authors evaluate the proposed method against current options in terms of performance and safety. The findings show that the suggested method is feasible and meets the necessary functional and security criteria.

Al Asad et al., (2021) [33] developed a permissioned blockchain using proof of authority (PoA) technology to provide confidentiality, permission sharing, and efficient distributed healthcare record administration. Furthermore, discussed the benefits of using de facto standards like Fast Health Interoperable Resources (FHIR) to facilitate the meaningful exchange of healthcare information among all parties involved and the significance of managing medical records in an interoperable manner. The study relies on simulated PoA use to learn about the performance of such a consensus process while using BT for safe information exchange.

Chelladurai et al., (2021) [34] established smart contracts on the blockchain to traditionally meet the needs of patients, doctors, and healthcare professionals. The proposed system introduces health models on the decentralized blockchain, such as immutable patient log creation using the Modified Merkle Tree data structure to store and quickly retrieve health records, medical record updates, health information exchange between providers, and viewership contracts. Many studies have been conducted to test the efficacy of the suggested system. The suggested system's performance regarding resources, transactions per second, and latency has been assessed using qualitative and quantitative indicators.

Abbas et al., (2023) [35] proposed the current Electronic Healthcare System and the primary constraints of current systems are their lack of scalability and privacy; thus, the purpose of this study is to define alternate approaches for achieving scalability, usability, and data protection in the healthcare system. It is essential in the healthcare field to provide complete, up-to-date patient records. Instant availability of patient records for improving efficiency and coordination is another quality that helps academics to think about effective EHR systems. The proposed model implementation makes use of blockchain-sharing technology, hyper-ledger protocols, and Proof-of-Authority.

Rathee et al., (2020) [36] establish a foundation for the safety of healthcare multimedia data using BT by producing a hash for each piece of information, allowing users throughout the network to be alerted to any unauthorized changes to data or medication. Simulation improvements are attributable to the usage of Blockchain have been used to confirm the findings, which show an 86% success rate in the face of product drop ratio, wormhole attack, falsification attack, and probabilistic authentication.

Anitha et al., (2023) [37] proposed Statistical Package for The Social Sciences (SPSS) to have a profound effect in many fields, including healthcare. The study of its potential uses in fields other than finance has been ongoing for some time. The author within the domain of health information technology is always struggling to stay up with the latest discoveries in the field because study in this area is still relatively new but increasing at a rapid rate. This paper includes findings from a survey of ongoing research into the potential medical applications of blockchain technology and an examination of the factors influencing the development of blockchain within the healthcare sector. The analysis reveals that multiple studies have proposed various applications for blockchain technology in the medical field. Even Nevertheless, there is a lack of prototypes implementations or studies that evaluate the efficacy of the stated use cases. The study expands our understanding of the current state of blockchain applications in the healthcare sector, including their potential benefits and limitations, and suggests directions for future investigation. Therefore, further study is required to fully comprehend, describe, and assess blockchain's potential benefits and drawbacks in the medical field. The primary objective of this study is to discover the elements that affect blockchain. Discriminant analysis and then path analysis has been employed in SPSS for analysis.

Gosh et al., (2023) [38] established the applications of this technology while also drawing attention to the challenges and potential areas for future blockchain study in the field of healthcare. An in-depth exploration of the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:29 UTC from IEEE Xplore. Restrictions apply.

blockchain and its capabilities serves as the paper's initial section. The remainder of the study is devoted to a comprehensive literature assessment of those articles chosen to represent the state of the art in a blockchain-based healthcare system research. Then, the answers offered by blockchain technology in healthcare systems and its primary areas of use are highlighted. The paper concludes with a discussion of its limitations, problems, and potential future research areas.

Lee et al., (2019) [39] introduce the SHARE Chain, a standardized architecture for sharing that includes two components designed to address interoperability and reliability concerns. First, it makes data more trustworthy by relying on a Blockchain registry's immutability; second, it establishes a Consortium Blockchain Network whereby only verified institutions can exchange information. The second enhancement is enhanced compatibility with healthcare data exchange standards through Quick Healthcare Interoperability Resources and Cross-Enterprise Document Sharing.

Wang et al., (2019) [40] suggest a novel method for exchanging individual health records that uses BT to prove the authenticity of the shared data. The novel system employs searching attribute-based encryption and symmetric encryption to safeguard privacy, keyword search, and fine-grained access control while exchanging personal health data. Moreover, the new scheme utilizes blockchain to maintain scheme keys, removing the potential for a single failure point associated with centralized key management. Lastly, the investigation of security threats and performance benchmarks confirm the safety and practicality of the method.

Wang et al., (2018) [41] employed attribute-based encryption (ABE)and Identity-based encryption (IBE) to protect sensitive patient information, while identity-based signatures (IBS) can be used to authenticate online transactions. Authors offer a new cryptographic primitive termed combined attribute-based/identity-based encryption and signature (C-AB/IB-ES) that combines the features of ABE, IBE, and IBS into a single system. It greatly reduces the complexity of system management and eliminates the requirement for additional cryptographic systems to provide different degrees of security. Table 3 indicates the comparison table of the literature of review.

TABLE III. COMPARISON OF LITERATURE OF REVIEW

Authors	Outcomes
Tan et al., (2022) [32]	The findings show that the suggested method is feasible and meets the necessary functional and security criteria.
Al Asad et al., (2021) [33]	It has been shown that proof of authority is superior to more conventional mining as a consensus method.
Chelladurai et al., (2021) [34]	The experimental findings demonstrate that BT reduces latency in the proposed network and increases throughput and performance using fewer resources.
Abbas et al., (2023) [35]	The proposed model implementation makes use of the blockchain sharing technology, hyper-ledger protocols, and Proof-of-Authority.
Rathee et al., (2020) [36]	The suggested technique outstrips the state-of-the-art by a factor of 86% in experimental analysis against unlawful actions carried out by malicious IoT devices such as the over products drop ratio, wormhole attack, falsification attack, and probabilistic authentication situations.
Anitha et al., (2023) [37]	The primary objective of this study is to discover the elements that affect blockchain. Discriminant analysis and then path analysis have been employed in SPSS for analysis.
Gosh et al., (2023) [38]	The solution offered by blockchain technology in healthcare systems and its primary areas of use are highlighted. The paper concludes with a discussion of its limitations, problems, and potential future research areas.
Lee et al., (2019) [39]	Compared and examined with comparable explore in various criteria (On/Off-chain, Identity, Interoperability, Patient-centered), where the suggested solutions meet all but one of the requirements.
Wang et al., (2019) [40]	The technique is secure and practically viable via performance assessment and security analysis.
Wang et al., (2018) [41]	The proposed models protect the privacy and authenticity of patient information. The authors conclude with a sample application in the health insurance industry.

III. PERFORMANCE EVALUTION

In this section, the author compares all the literature survey papers based on EHR encryption efficiency comparison and feature comparison.

A. EHR Encryption efficiency comparison

The time required for encryption is compared in Figure 4 for the [40] and [41] systems as well as the [32] scheme. The y-axis displays the encryption time in seconds (s), while the x-axis displays the file size in kilobytes (KB). To ensure that

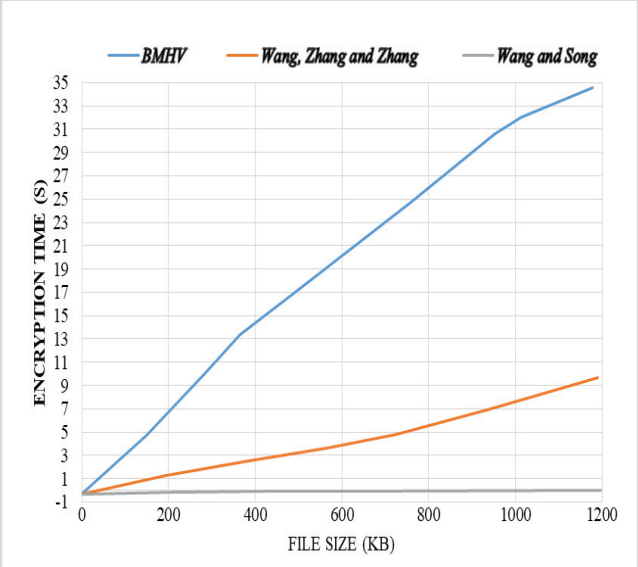


Fig. 4. Encryption efficiency comparison [32]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:29 UTC from IEEE Xplore. Restrictions apply.

the analysis of the exact file sizes utilized in the two approaches above is accurate, the size of the file subject used for the comparison is restricted to 1200 kilobytes (1.2 MB). The results show that BHMV's encryption algorithm is more effective than the alternatives studied. In other words, the encryption time scales better with file size than the competing approaches. BMHV's suggested method is future-proof and feasible, providing a quicker overall encryption performance that can be used in the system's massive EHR storage and sharing infrastructure.

B. Features comparison of schemes

Features including server-side verifiability, encryption, KW search, user-side verifiability, storage integrity, and dynamic updating are all compared and contrasted in Table 4. These systems rely entirely on blockchain for security, which is inadequate for keeping information private. Comparing the two, it is clear that author [32] supports all the essential characteristics, whereas author [33] supports all the features apart from storage integrity. Similarly, the author supports all major features [40] except dynamic updates.

TABLE IV. MAJOR ASPECTS OF BLOCKCHAIN-BASED HEALTHCARE MANAGEMENT SYSTEMS (SUPPORTED = ✓; UNSUPPORTED = ✗).

Authors	Encryption	KW Search	Server-side Verifiability	User-side Verifiability	Storage Integrity	Dynamic Update
Wang et al., (2018) [41]	✓	✓	✗	✗	✓	✗
Wang et al., (2019) [40]	✓	✓	✓	✓	✓	✗
Lee et al., (2019) [39]	✗	✓	✗	✗	✓	✗
Al Asad et al., (2021) [33]	✗	✗	✗	✗	✓	✗
Chelladurai et al., (2021) [34]	✗	✗	✗	✗	✓	✓
Tan et al., (2022) [32]	✓	✓	✓	✓	✓	✓

IV. CONCLUSION AND FUTURE SCOPE

An EHR is software that stores and organizes a person's medical history. Several medical facilities' records are stored in its database. Emerging technologies like blockchain are being used to provide novel solutions across industries, including healthcare. The healthcare industry uses a blockchain to store and transfer patient data across labs, pharmacies, clinics, hospitals, and doctors' offices.

The medical industry is one where BT can be employed to detect serious and perhaps life-threatening errors properly. In this study, authors examine the use of BT for managing EHR. Medical facilities can benefit from this technology since it facilitates a deeper study of patient information. After an examination of the relevant literature, it draws parallels between different studies. Encryption, KW search, server and user verifiability, storage integrity, and dynamic updates are all features that have been compared and contrasted. Comparing the two, it is clear that author [32] supports all the essential characteristics, whereas author [33] supports all the features apart from storage integrity. Similarly, the author supports all major features [40] except dynamic updates.

Important for patients and necessary for any subsequent treatment, particular rules for the electronic storage of patient data, medical supplies, equipment, and materials utilized, etc., should be created in the future. Then, it's necessary to specify which data can be encrypted and stored publicly through permissionless blockchains and which can be encrypted and stored privately, with access restricted to approved network members.

REFERENCES

[1] P. P. Ray, Di. Dash, K. Salah, and N. Kumar, "Blockchain for IoT-Based Healthcare: Background, Consensus, Platforms, and Use Cases," *IEEE Syst J*, vol. 15, no. 1, pp. 85–94, Mar. 2021, doi: 10.1109/JSYST.2020.2963840.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] A. K. Yadav, Shweta, and D. Kumar, "Blockchain technology and vaccine supply chain: Exploration and analysis of the adoption barriers in the Indian context," *Int J Prod Econ*, vol. 255, p. 108716, Jan. 2023, doi: 10.1016/J.IJPE.2022.108716.

[3] C. De Pietro and I. Francetic, "E-health in Switzerland: The laborious adoption of the federal law on electronic health records (EHR) and health information exchange (HIE) networks," *Health Policy (New York)*, vol. 122, no. 2, pp. 69–74, Feb. 2018, doi: 10.1016/J.HEALTHPOL.2017.11.005.

[4] H. Singh *et al.*, "Primary care practitioners' views on test result management in EHR-enabled health systems: a national survey," *Journal of the American Medical Informatics Association*, vol. 20, no. 4, pp. 727–735, Jul. 2013, doi: 10.1136/AMIAJNL-2012-001267.

[5] A. F. Klaib and M. S. Nuser, "Evaluating EHR and health care in Jordan according to the international health metrics network (HMN) framework and standards: A case study of hakeem," *IEEE Access*, vol. 7, pp. 51457–51465, 2019, doi: 10.1109/ACCESS.2019.2911684.

[6] M. Alarjani, M. Alhaider, and H. F. Ahmad, "A Review of Challenges of Block Chain with COVID-19: A Review Paper," *European Journal of Health Sciences*, vol. 8, no. 2, pp. 32–49, Mar. 2023, doi: 10.47672/EJHS.1384.

[7] X. Li *et al.*, "A Quantitative and Qualitative Review of Blockchain Research from 2015 to 2021," *Sustainability*, vol. 15, no. 6, p. 5067, Mar. 2023, doi: 10.3390/su15065067.

[8] R. Natarajan, G. H. Lokesh, F. Flammini, A. Premkumar, V. K. Venkatesan, and S. K. Gupta, "A Novel Framework on Security and Energy Enhancement Based on Internet of Medical Things for Healthcare 5.0," *Infrastructures (Basel)*, vol. 8, no. 2, Feb. 2023, doi: 10.3390/infrastructures8020022.

[9] C. Wachira *et al.*, "Analysis of user interactions with a digital health wallet for enabling care continuity in the context of an ongoing pandemic," *Journal of the American Medical Informatics Association*, vol. 30, no. 4, pp. 674–682, Mar. 2023, doi: 10.1093/JAMIA/OCAD004.

[10] N. Islam, Y. Faheem, I. U. Din, M. Talha, M. Guizani, and M. Khalil, "A blockchain-based fog computing framework for activity recognition as an application to e-Healthcare services," *Future Generation Computer Systems*, vol. 100, pp. 569–578, Nov. 2019, doi: 10.1016/j.future.2019.05.059.

[11] E. Chukwu and L. Garg, "A systematic review of blockchain in healthcare: Frameworks, prototypes, and implementations," *IEEE Access*, vol. 8, pp. 21196–21214, 2020, doi: 10.1109/ACCESS.2020.2969881.

[12] T. Ali Syed, A. Alzahrani, S. Jan, M. S. Siddiqui, A. Nadeem, and T. Alghamdi, "A Comparative Analysis of Blockchain Architecture and its Applications: Problems and Recommendations," *IEEE Access*, vol. 7, pp. 176838–176869, 2019, doi: 10.1109/ACCESS.2019.2957660.

[13] C. Esposito, A. De Santis, G. Tortora, H. Chang, and K.-K. Raymond Choo, "DEPARTMENT: Cloud and the Law Blockchain: A Panacea for Healthcare Cloud-Based Data Security and Privacy?" [Online]. Available: www.computer.org/cloud

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:29 UTC from IEEE Xplore. Restrictions apply.

[14] A. Al Omar, M. Z. A. Bhuiyan, A. Basu, S. Kiyomoto, and M. S. Rahman, "Privacy-friendly platform for healthcare data in cloud based on blockchain environment," *Future Generation Computer Systems*, vol. 95, pp. 511–521, Jun. 2019, doi: 10.1016/j.future.2018.12.044.

[15] K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," *IEEE Access*, vol. 4. Institute of Electrical and Electronics Engineers Inc., pp. 2292–2303, 2016. doi: 10.1109/ACCESS.2016.2566339.

[16] A. Hasselgren, K. Kralevska, D. Gligoroski, S. A. Pedersen, and A. Faxvaag, "Blockchain in healthcare and health sciences—A scoping review," *International Journal of Medical Informatics*, vol. 134. Elsevier Ireland Ltd, Feb. 01, 2020. doi: 10.1016/j.ijmedinf.2019.104040.

[17] J. Xie *et al.*, "A Survey of Blockchain Technology Applied to Smart Cities: Research Issues and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 3, pp. 2794–2830, Jul. 2019, doi: 10.1109/COMST.2019.2899617.

[18] A. Farouk, A. Alahmadi, S. Ghose, and A. Mashatan, "Blockchain Platform for Industrial Healthcare: Vision and Future Opportunities."

[19] T. Hardin and D. Kotz, "Blockchain in health data systems: a survey."

[20] I. Yaqoob, K. Salah, R. Jayaraman, and Y. Al-Hammadi, "Blockchain for healthcare data management: opportunities, challenges, and future recommendations," *Neural Comput Appl*, vol. 34, no. 14, pp. 11475–11490, Jul. 2022, doi: 10.1007/s00521-020-05519-w.

[21] S. Namasudra, G. Chandra, and D. Editors, "Studies in Big Data 83." [Online]. Available: http://www.springer.com/series/11970.

[22] B. "Didi" Davis and A. Swenson, "How Carequality, The Sequoia Project, and eHealth Exchange Support the Interoperable Exchange of Health Data in the USA," *J Digit Imaging*, vol. 35, no. 4, pp. 812–816, Aug. 2022, doi: 10.1007/s10278-021-00538-y.

[23] D. Evans, R. McDonald, and T. Coatta, "Access controls and healthcare records: Who owns the data?," *Communications of the ACM*, vol. 62, no. 7. Association for Computing Machinery, pp. 41–46, Jul. 01, 2019. doi: 10.1145/3331142.

[24] D. V. Dimitrov, "Blockchain applications for healthcare data management," *Healthc Inform Res*, vol. 25, no. 1, pp. 51–56, 2019, doi: 10.4258/hir.2019.25.1.51.

[25] E. Hossain et al., "Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review," Computers in Biology and Medicine, vol. 155. Elsevier Ltd, Mar. 01, 2023. doi: 10.1016/j.compbiomed.2023.106649.

[26] S. Khezr, M. Moniruzzaman, A. Yassine, and R. Benlamri, "Blockchain technology in healthcare: A comprehensive review and directions for future research," *Applied Sciences (Switzerland)*, vol. 9, no. 9, May 2019, doi: 10.3390/app9091736.

[27] A. A. Siyal, A. Z. Junejo, M. Zawish, K. Ahmed, A. Khalil, and G. Soursou, "Applications of blockchain technology in medicine and healthcare: Challenges and future perspectives," *Cryptography*, vol. 3, no. 1, pp. 1–16, Mar. 2019, doi: 10.3390/cryptography3010003.

[28] H. T. Vo, A. Kundu, and M. Mohania, "Research directions in blockchain data management and analytics," in *Advances in Database Technology - EDBT*, OpenProceedings.org, 2018, pp. 445–448. doi: 10.5441/002/edbt.2018.43

[29] S. J. Wang *et al.*, "A cost-benefit analysis of electronic medical records in primary care," *American Journal of Medicine*, vol. 114, no. 5, pp. 397–403, Apr. 2003, doi: 10.1016/S0002-9343(03)00057-3.

[30] R. H. Miller and I. Sim, "Physicians' use of electronic medical records: Barriers and solutions," *Health Aff*, vol. 23, no. 2, pp. 116–126, 2004, doi: 10.1377/hlthaff.23.2.116.

[31] A. L. Terry *et al.*, "Implementing electronic health records Key factors in primary care Implantation du dossier de santé électronique Principaux facteurs pour les soins de première ligne," 2008.

[32] T. L. Tan, I. Salam, and M. Singh, "Blockchain-based healthcare management system with two-side verifiability," *PLoS One*, vol. 17, no. 4, p. e0266916, Apr. 2022, doi: 10.1371/JOURNAL.PONE.0266916.

[33] A. Al Hussain, M. A. Emon, T. A. Tanna, R. I. Emon, and M. M. H. Onik, "A Systematic Literature Review of Blockchain Technology Adoption in Bangladesh," *Annals of Emerging Technologies in Computing*, vol. 6, no. 1. International Association for Educators and Researchers (IAER), pp. 1–30, 2022. doi: 10.33166/AETiC.2022.01.001.

[34] U. Chelladurai and S. Pandian, "A novel blockchain based electronic health record automation system for healthcare," *J Ambient Intell Humaniz Comput*, vol. 13, no. 1, pp. 693–703, Jan. 2022, doi: 10.1007/s12652-021-03163-3.

[35] A. Abbas and Md. A. Hamid, "Adapting hybrid approaches for electronic medical record management and sharing using blockchain sharding," *Periodicals of Engineering and Natural Sciences*, vol. 11, no. 1, pp. 5–14, Jan. 2023, doi: 10.21533/PEN.V11I1.3405.

[36] G. Rathee, A. Sharma, H. Saini, R. Kumar, and R. Iqbal, "A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology," *Multimed Tools Appl*, vol. 79, no. 15–16, pp. 9711–9733, Apr. 2020, doi: 10.1007/s11042-019-07835-3.

[37] C. Anitha, A. Joshi, V. Thumar, K. M. J, and P. Viswanath, "Analyzing the Factors Influencing in Blockchain Innovation in Health Care," *Journal for ReAttach Therapy and Developmental Diversities*, vol. 6, no. 2s, pp. 330–337, Mar. 2023, Accessed: Apr. 15, 2023. [Online]. Available: https://www.jrtdd.com/index.php/journal/article/view/302

[38] P. K. Ghosh, A. Chakraborty, M. Hasan, K. Rashid, and A. H. Siddique, "Blockchain Application in Healthcare Systems: A Review," *Systems 2023, Vol. 11, Page 38*, vol. 11, no. 1, p. 38, Jan. 2023, doi: 10.3390/SYSTEMS11010038.

[39] I. Yoo, J. Bi, X. Hu, National Science Foundation (U.S.), and Institute of Electrical and Electronics Engineers, Proceedings, 2019 IEEE International Conference on Bioinformatics and Biomedicine : November 18-21, 2019, San Diego, CA, USA.

[40] Y. Liu *et al.*, "A Blockchain-Based Personal Health Record System for Emergency Situation," *Security and Communication Networks*, vol. 2022, 2022, doi: 10.1155/2022/4941214.

[41] H. Wang and Y. Song, "Secure Cloud-Based EHR System Using Attribute-Based Cryptosystem and Blockchain," *J Med Syst*, vol. 42, no. 8, Aug. 2018, doi: 10.1007/s10916-018-0994-6.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:20:29 UTC from IEEE Xplore. Restrictions apply.

Capacitor Trapped Current Fed ZCS-ZVS Push Pull DC-DC converter for renewable energy conversion

Dhananjay Yadav, Vishal Verma
Electrical Engineering Department
Delhi Technological University
Shahbad Daulatpur Village, Rohini, New Delhi, Delhi 110042
yadavdhananjay552@gmail.com vishalverma@dce.ac.in

2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE) | 979-8-3503-1312-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICEPE57949.2023.10201575

***Abstract*— A current fed ZCS-ZVS push-pull converter has been analyzed and designed for renewable energy conversion applications. The topology maintains both zero current switching (ZCS) and (ZVS) on leading and trailing edges owing to the resonance between the trapped capacitor and the leakage inductance of the HF transformer, resulting in a near sinusoidal waveform at the primary of the HF transformer. The voltage spike across the power devices are suppressed effectively enabling the selection of low-voltage, low cost power devices. The proposed converter no auxiliary circuit is involved and since all the switches are ground referred, non-isolated driver with single power supply can be employed. A comprehensive study of 500 W highly efficient push pull converter and design of the proposed converter including steady-state and transient analysis has been reported.**

Keywords—Push Pull, ZCS, ZVS, Trapped Capacitor, Current fed DC-DC converter

I. INTRODUCTION

In light of the scarcity of fossil fuels and concerns regarding climate change, interest in renewable energy has grown considerably in recent years. As a sustainable way to satisfy our energy demands, renewable energy sources, help lowering of greenhouse gas emissions thereby reducing the effects of climate change. Solar energy being abundant and closest to the loading point, but struggle with intermittency issues and generally yield low output voltages, and thus require step-up dc-dc converters having high conversion ratio with low input current ripple for its effective utilization for off grid and grid connected application [1][8]. High efficiency, low EMI, low weight, lower cost, high power density, extended lifetime, among are major factors for design considerations of charging infrastructure for EV loads for off grid applications, where charging station is required to be supported by PV fed generation to charge intermediate batteries[9].

DC-DC converters can be either isolated or non-isolated, however, the isolated converters with High-frequency (HF) transformer are chosen over non-isolated topologies because of their advantages in galvanic isolation, low ripple contents, enhanced regulation, flexible output voltage, and provide multiple output at different levels with ease. Electrical isolation between the input and output reduces the risk of electrical shock and protect the delicate components against ground events, and electrical noise.

Current-fed converters are considered front runner in various step-up dc-dc converters applicable to PV fed system. These converters naturally come with smaller input current pulsations and due to pre-boosting paving the advantage of a lower transformer turns ratio compared to voltage-fed converters with output voltage clamping, risks of HF transformer getting saturated reduced largely and there are no duty cycle loss issues. Therefore, for low-voltage input renewable energy conversion applications, current-fed DC–DC converters are preferable over voltage-fed versions [2].

The most common topologies for isolated current-fed step-up DC-DC converters employ flyback, half-bridge, full-bridge, and push-pull (PP) architectures. For galvanic isolation initially line frequency (50–60 Hz) transformers [3] are employed, although dependable and simple, its large size and weight is major impediment. Flyback converter falter on account of large losses and significant voltage drop as the step up voltage output increases. On the other hand; compared to the PP or full-bridge converter configuration, the half-bridge converter configuration falls short for high step up application since it requires twice the turn ratio of the comparable high-frequency transformer. Features like simple circuitry, fewer power switches, reduced conduction losses, high step up voltage conversion ratio, better transformer utilization makes the PP converter more suitable as compared to full bridge configurations. Owing to utilization of the leakage inductor of the HF transformer as integrated magnetic solution, the power density is increased drastically for energy transfer in current-fed topologies [3].

Traditional current-fed push-pull converters have a number of disadvantages, including high-voltage spikes across switches resulting from the resonance between the switching loop parasitic capacitor and the leakage inductance of the HF transformer, and high switching losses due to the hard-switching operation of the primary switches. Passive snubbers, energy recovery circuits, and active clamp snubbers are reported in the literature [3] to provide protection against high spikes on the switches. Active-clamp snubbers are favourably utilized as their auxiliary circuit may not only suppressed the voltage spikes but also assisted in zero voltage switching (ZVS) thereby showing high efficiency hence, paving the way for increase in the switching frequency and reduction of the converter size[4]. For further effective remediation, numerous enhanced topologies with zero-voltage switching (ZVS) [3], zero–current switching (ZCS)[4], and voltage spike clamping and suppression [4] are suggested in literature. The resonant push-pull converters, which use the circuit parasitic, or resonant tank to help with zero-voltage switching (ZVS) or zero-current switching (ZCS) for primary switches, is also reported [5].These methods achieve a high conversion efficiency while reducing the switching loss drastically.

A new simple active-clamped ZCS-ZVS current-fed push–pull DC–DC converter for PV applications is proposed. The proposed converter includes a current-fed push–pull circuit with a trapped clamping capacitor, one HF transformer, a rectifier with fast-recovery SIC diodes and a capacitive filter to output power to 380V DC bus. Both ZCS and ZVS on leading and trailing edges are attained owing to the resonance

979-8-3503-1312-3/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:55:16 UTC from IEEE Xplore. Restrictions apply.

between the trapped capacitor and the leakage inductance of the HF transformer, resulting in a near sinusoidal waveform at the primary of the HF transformer. The voltage spike across the power devices are suppressed effectively enabling the selection of low-voltage, low cost power devices. In the proposed converter no auxiliary circuit is involved and since all the switches are ground referred, non-isolated driver with single power supply can be employed with very simple control. The power range of the proposed converter is 3 KW. A 500 W scaled down model was implemented in this paper. Figure 1 Block diagram of the architecture of the renewable energy conversion system. The paper briefly analysis and design the key characteristics and simulation results are discussed thereby demonstrating the capabilities of the converter in terms of boosting the voltage with desired power at high efficiency while undergoing ZCS-ZVS operations.

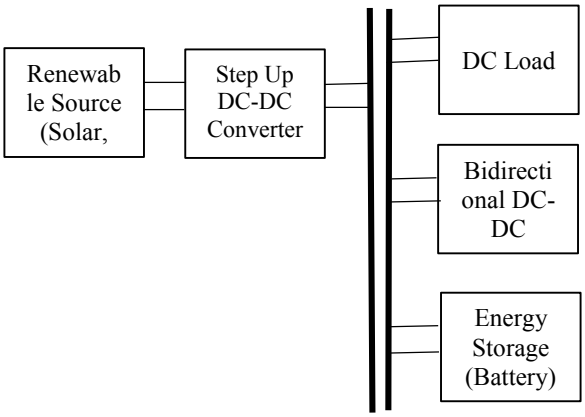


Fig. 1 Architecture of renewable energy conversion system

II. System Configuration

The proposed converter configuration employs an active clamped ZCS-ZVS operation of the 500W current fed push pull converter as shown in Figure 2. The converter operates on low voltage DC bus duly fed by MPPT charge controller having voltage of 48V. The new configuration employ 2 blocking diodes with a trapped clamping capacitor on the primary side of conventional push pull converter which enable the configuration to achieve ZCS-ZVS operation of the switches. The output side of the HF transformer is assumed connected with fast recovery SIC diodes to formulate a DC bus having output voltage 380V with a capacitive filter.

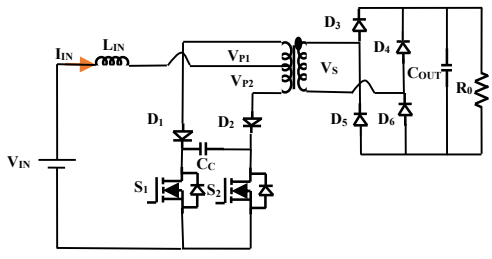


Fig. 2 Schematic of Proposed Converter

III. Modes of Operation

The proposed naturally clamped ZCS-ZVS Push Pull converter's steady-state operation and analysis, together with pertinent mathematical equations and assumptions, is summarized in this section. The resonant tank for the ZCS functioning of the primary switches is composed of an trapped capacitor C_C and transformer leakage inductance L_{Leakge}. The gating signals of the switches have a duty cycle D that is larger than 50% and are 180 °phase-shifted. Figure 3 shows the waveforms under steady conditions. The analysis was completed for the first half cycle, the intervals were repeated for the second half cycle with the other symmetrical devices conducting. Figure 4 shows analogous circuits that represent the various operating modes of the converter.

The quantities and ratings of components as well as the theoretical converter performance can be determined using analytical equations that were generated.

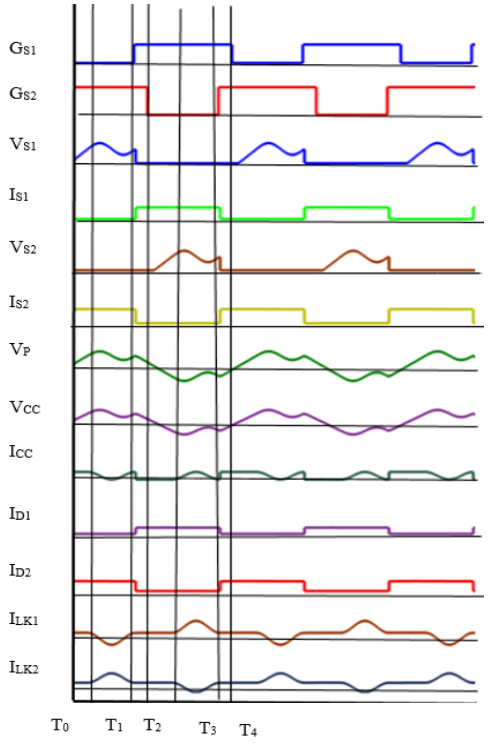


Fig. 3 Switching pulses and steady state waveforms

A. Interval 1 (Fig. 4(a): T0< t < T1')

In this interval the primary side switch S_2 is conducting and diode D_3 and D_6 of secondary side Rectifier Bridge are conducting. Constant current I_{IN} flows through $L_{Leakage1}$ and $L_{Lekage2}$ .Switch S_2 carries the entire input current. The trapped capacitor's voltage ($V_{C\ trapped}$) is clamped at $2V_{P1}$.

$$V_{P1} = V_{P2}$$

$$V_{IN} = V_L + V_{P1}$$
$$V_{IN} = V_L - V_{P2} + V_{CC}$$
$$V_{P1} = \frac{V_{C\ trapped}}{2}$$
$$V_L = V_{IN} - \frac{V_{c\ trapped}}{2}$$

$$V_{C\ trapped} = 2V_{P1}$$

B. Interval 2 (Fig. 4(a): T1' < t < T1)

The clamping capacitor starts discharging through the path: Clamping capacitor via Diode D_1 via S_1 parasitic capacitor to achieve ZCS and ZVS. Current through $L_{Leakage1}$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:55:16 UTC from IEEE Xplore. Restrictions apply.

starts decreasing sinusoidaly and current through $L_{Leakage2}$ starts increasing sinusoidaly.

C. *Interval 3 (Fig. 4(b): T1< t < T2)*

Switch S_1 is turned ON and all the primary side switches conduct simultaneously during this interval. The current through S_2 decreases and current through S_1 increases. The input current gets distributed in both primary leakage inductors due to the presence of clamping capacitor. Polarity across the capacitor got interchanged which signifies that the capacitor starts discharging which is supported by previous mode. Therefore ZCS-ZVS for switch S1 (turning ON) and S2 (turning OFF) gets achieved.

D. *Interval 4 (Fig. 4(c): T2< t <T2')*

At T_2 the switch S_2 gets turned OFF and only S_1 is conducting diode D_5 and D_4 of secondary side Rectifier Bridge are conducting. Constant current Iin flows through $L_{Leakage1}$ and $L_{Leakage2}$ .Switch S_2 carries the entire input current. The trapped capacitor's voltage ($V_{C\ trapped}$) is clamped at $-2V_{P1}$.

E. *Interval 5 (Fig. 4(c): T2' <t<T3)*

At T2' current through $L_{Leakage1}$ starts increasing sinusoidaly and current through $L_{Leakage2}$ starts decreasing sinusoidaly and trapped capacitor Cc starts charging until it reaches its maximum value.

$$V_{P1} = V_{P2}$$
$$V_{IN} = -V_L - V_{P1} + V_{CC}$$
$$V_{P1} = \frac{V_{C\ trapped}}{2}$$
$$V_L = \frac{V_{c\ trapped}}{2} - V_{IN}$$
$$V_{C\ trapped} = 2V_{P1}$$

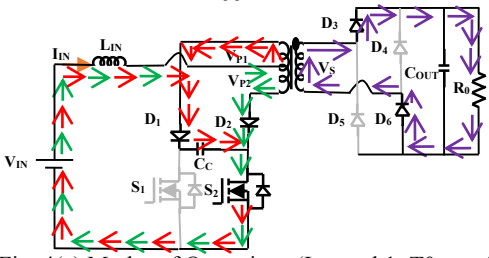


Fig. 4(a) Modes of Operations (Interval 1: T0< t < T1)

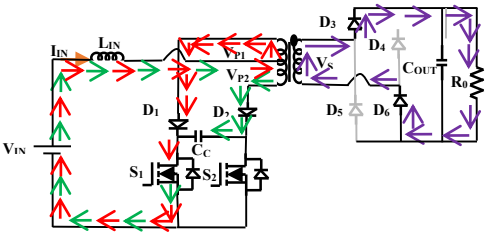


Fig. 4(b) Modes of Operations (Interval 1: T1< t < T2)

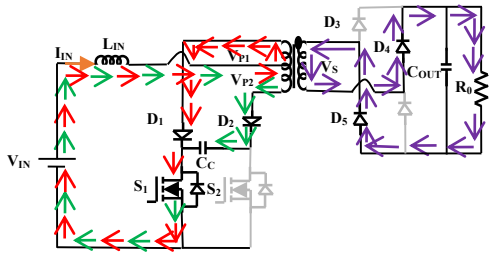


Fig. 4(c) Modes of Operations (Interval 1: T2< t < T3)

IV. System Analysis

This work explains the analysis of the current fed push-pull converter shown in Figure 2. The notable characteristics of this converter include high efficiency, ZVS ,ZCS and a low switching surge. In addition, the inductor's series connection with the input power source results in the suppression of backward noise and high immunity against external noise. Two operating modes are analysed in this study: the first when the switch S_1 or S_2 is ON, and the other when both switches S_1 and S_2 are ON. Second, the effects of the transformer's leakage inductance and trapped capacitor were discussed in detail. The formulation of differential equations that reflect the low-frequency behaviour of the resonant converter is the initial step in the analytical process. Differential equations are derived in this analysis, and a few analytical expressions are explained. Calculations of the static and dynamic properties are also discussed. Our mathematical model considers the effects of the inductance of the transformer and trapped capacitor. Resistances against parasites were disregarded. Diodes are considered ideal for the analysis.

$V_{IN}(t)$: Input Voltage
$V_{GS1}(t), V_{GS2}(t)$: Gate to Source Voltage of Mosfet Switch S_1 and S_2 respectively
$V_{C\ trapped}(t)$: Voltage across trapped capacitor
$I_{IN}(t)$: Input inductor current
$I_{OUT}(t)$: Output current
$V_{OUT}(t)$: Output Voltage
L_{TF}: Total leakage inductance of transformer's primary

A. *State 1:Either S_1 or S_2 is on.*

The circuit equations in this interval are:

$$V_{IN} = L\frac{di}{dt} + \frac{d\emptyset}{dt} \quad (1)$$

Where ∅ is linkage flux associated with primary winding of the transformer

$$I_{C\ trapped} = C_{trapped}\frac{dV_{C\ trapped}}{dt} \quad (2)$$

$$(I_{IN} - I_{C\ trapped}) - I_{C\ trapped} - nI_{OUT} = \frac{\emptyset}{L_{TF}} \quad (3)$$

$$V_{C\ trapped} = 2\frac{d\emptyset}{dt} \quad (4)$$

$$C_{OUT}\frac{dV_{OUT}}{dt} + \frac{V_{OUT}}{R} = I_{OUT} \quad (5)$$

Where n is turns ratio of transformer $n = \frac{N_2}{N_1}$
Equation (3) represents the relation of transformers magnetizing current and linkage flux.
By using Laplace transform technique to solve (1) to (5) we get:

$$\emptyset(t) = \emptyset(0) + A\left(t - \frac{\sin w_0 t}{w_0}\right) + \frac{B}{w_0}(1 - \cos w_0 t)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:55:16 UTC from IEEE Xplore. Restrictions apply.

$$I_{IN}(t) = I_{IN}(0) + \frac{V_{IN}(t)}{L} - \frac{A}{L}\left(t - \frac{\sin w_0 t}{w_0}\right) - \frac{B}{w_0 L}(1 - \cos w_0 t)$$

$$V_{OUT}(t) = V_{OUT}(0)\left(1 - \frac{t}{C_{OUT}R} - \frac{t^2}{2C_{OUT}}\right) + \frac{I_{OUT}(0)t}{C_{OUT}} - \frac{nAC_{OUT}R^2}{1 + (w_0 C_{OUT}R)^2}\left[1 - \frac{t}{C_{OUT}R} - \frac{{w_0}^2 t^2}{2} - \cos w_0 t + \frac{\sin w_0 t}{w_0 C_{OUT}R}\right] + \frac{nBC_{OUT}R^2}{1 + (w_0 C_{OUT}R)^2}\left[\frac{1}{w_0 C_{OUT}R} + w_0 t - \frac{\cos w_0 t}{w_0 C_{OUT}R} - \sin w_0 t\right]$$

$$V_{C\ trapped}(t) = 2A(1 - \cos w_0 t) + 2B \sin w_0 t \qquad (6)$$

Where, $\emptyset(0)$, $I_{IN}(0)$, $V_{OUT}(0)$, $I_{OUT}(0)$ are initial values in this interval and the resonance frequency w_0 and parameters A and B are:

$$w_0 = \frac{1}{2}\sqrt{\frac{1}{C_{trapped}}\left(\frac{1}{L} + \frac{1}{L_{TF}}\right)}$$

$$A = \frac{V_{IN} + nV_{OUT}L}{4{w_0}^2 L C_{trapped}}$$

$$B = \frac{I_{IN}(0) - nI_{OUT}(0) - \frac{\emptyset(0)}{L_{TF}}}{4w_0 C_{trapped}}$$

This state shifts to next state when the resonance voltage $V_{C\ trapped}$ reaches to zero.

B. State 2: When S_1 and S_2 are on.

The circuit equations in this interval are:

$$V_{IN} = L\frac{di}{dt}$$

$$C_{OUT}\frac{dV_{OUT}}{dt} + \frac{V_{OUT}}{R} = 0$$

Using Laplace transform

$$\emptyset(t) = \emptyset(t1)$$

$$I_{IN}(t) = \frac{V_{IN}}{L}(t - t_1)$$

$$V_{OUT}(t) = V_{OUT}(t_1)\left(1 - \frac{t - t_1}{C_{OUT}R}\right)$$

Where t_1 is the time when both switches S_1 and S_2 are ON Putting the differentiation in the above differential equations and to be zero static characteristics are derived.

$$2\emptyset = A\left(t - \frac{\sin w_0 t}{w_0}\right) + \frac{B}{w_0}(1 - \cos w_0 t)$$

$$V_{IN}T_S = A\left(t - \frac{\sin w_0 t}{w_0}\right) + \frac{B}{w_0}(1 - \cos w_0 t)$$

$$V_{OUT}T_S = nA\left(t - \frac{\sin w_0 t}{w_0}\right) + \frac{nB}{w_0}(1 - \cos w_0 t)$$

Where $\emptyset$ is linkage is flux and T_S is the switching period of the converter operation

Above equations yields the following results:

$$|\emptyset| = \frac{1}{2}V_{IN}T_S$$

$$V_{OUT} = nV_{IN}$$

Maximum value $V_{C\ trapped\ ,max}$ of the resonance voltage is derived from (6) as follows:

$$V_{C\ trapped\ ,max} = 2A + 2\sqrt{A^2 + B^2}$$

V. PERFORMANCE ANALYSIS

The circuit of the proposed active clamped ZCS-ZVS push pull converter is implemented in MATLAB Simulink to verify the theoretical analysis. The converter is tested with a 48V DC input voltage with a 100 kHz switching frequency. 380V and 500W are the specified output voltage and power, respectively. The Duty D at the specified input voltage is obtained as 0.58. The leakage inductance $L_{Leakage1}$ and $L_{Leakage2}$ of the transformer T, which has no air gaps, is 2 µH. By using a boost inductor L_{IN} of 500 µH, the maximum ripple current ΔI_{IN}, max is set to 0.3 A at D 0.58. Two blocking diodes with a trapped capacitor C_C has been added on the primary side of conventional push pull converter to achieve ZCS- ZVS. The voltage across the trapped capacitor is chosen from the circuit analysis presented in section IV above. The critical simulation parameters are listed in the Table I.

TABLE I. SIMULATION PARAMETERS

Parameter	Values
Input voltage	48 V
Output voltage	380 V
Output power	500 W
Switching frequency	100 kHz
Turns Ratio	1: 1: 4.3
Leakage inductance ($L_{Leakage1}$and $L_{Leakage2}$)	2 μH
Inductance of boost inductor	500 μH
Trapped capacitor	50 nF
Output capacitor	470 μF

Figure 5 displays the output current and voltage waveforms measured at a 48 V input and a 500 W load condition. Figure 5 shows that due to the presence of capacitive output filter the maximum output voltage ripple is 0.5V.Due to the boost inductor's presence at the input, the maximum ripple current seen in Figure 6 input current waveform is restricted to 10.04 A. In addition to the gating signals for switches S_1 and S_2, Figure 7 shows the waveforms of the current and drain-source voltage for S_1 and S_2, respectively. Both ZCS-ZVS are obtained for switches S_1 and S_2, respectively, as shown in Figure 7 which eliminates the switching conduction losses hence increasing efficiency. Current is at 0 A before switches S_1, S_2 turn off, ensuring the ZCS turn off of both switches. Figure 8, respectively, shows the waveforms of the trapped capacitor voltage and the primary voltages (V_{P1} and V_{P2}) of the transformer. Resonance on the primary side of the transformer causes a voltage

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:55:16 UTC from IEEE Xplore. Restrictions apply.

waveform with a sinusoidal shape at the primary. It is clear that the trapped capacitor voltage is double the primary voltages of the transformer, as determined in the analytical section. As a result of the inclusion of quick recovery SIC diodes on the output side, the secondary voltage waveform of the transformer in Figure 9 is more like a sinusoidal waveform.

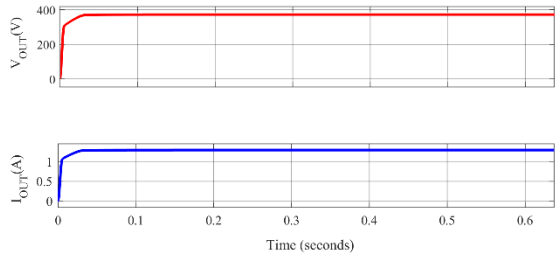


Fig. 5 Output Current $I_{OUT}(A)$ and Output Voltage $V_{OUT}(V)$

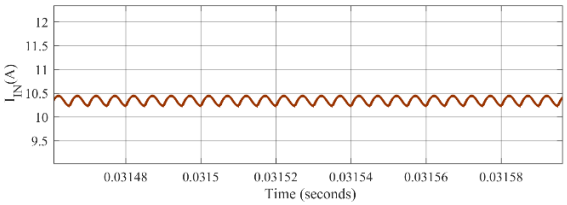


Fig. 6 Input Current $I_{IN}(A)$

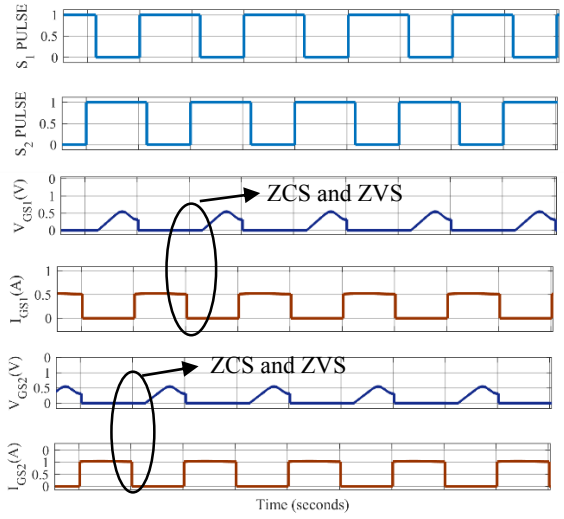


Fig. 7 $S_1 and\ S_2$ gate pulses with drain to source voltage and current

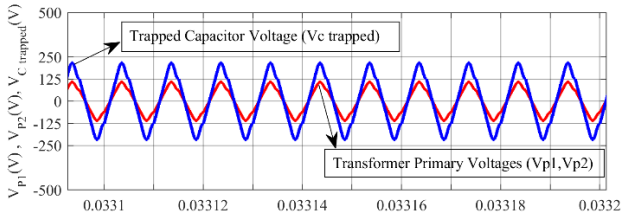


Fig. 8 Transformer primary voltages $V_{P1}\ (V) and\ V_{P2}(V)$ with Trapped Capacitor voltage $V_{C\ trapped}(V)$

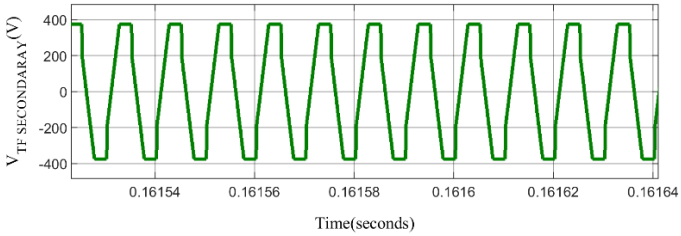


Fig. 9 Transformer secondary voltage $V_{TF\ SECONDARY}(V)$

A. Comparison of Proposed Converter and the Existing Topologies

Table II compares the efficiency, primary voltage waveform shape, clamping state, soft-switching, and quantity of ground-referred switches and gate drivers of three distinct Push Pull converter topologies [6][7]. The suggested converter contains fewer active power devices than the topologies in [6] and [7], as shown. The proposed converter has the maximum efficiency when compared to the converters in [6] and [7].The resonance at the primary side of the transformer causes the primary voltage to be sinusoidal, unlike in these two converters .As all switches are ground referenced, the circuit design is made simpler and fewer isolation gate drivers are needed. As compared to topology in [7] number of clamping capacitor required is less. In contrast to existing topologies, the suggested converter can provide ZCS-ZVS both operation for both switches that are present at primary.

TABLE II. COMPARITIVE ANALYSIS

Parameters	**Topology**		
	Conventional Push Pull [6]	Modified Push Pull [7]	Proposed Push Pull
Efficiency	High	Higher	Highest
Transformer Primary Voltage	Square	Quasi square	Sinusoidal
Clamping	NO	YES	YES
ZVS/ZCS	NO	ZCS	BOTH
Ground Referred Switches	Yes	NO	Yes
Gate Driver Required	1	2	1
Clamping Capacitor Required	0	2	1

The switch current and drain to source voltage waveforms of the three topologies are compared in Figure 10. Along with high voltage conversion ratio, the voltage stress on the switches is low due to the ZVS and ZCS operations of both the primary side switches as shown in figure 10.The proposed converter has no voltage spike during switching as compared to topologies presented in [6] and [7] as shown in Figure 10. Also switch drain to source voltage in the presented converter is sinusoidal in shape as compared to other two topologies [6][7]. Figure 12 represents the comparison between transformer's primary voltages for all three topologies. The proposed converter features a sinusoidal waveform at the transformer primary, which results in reduced losses in the transformer and less leakage flux, further leading to greater converter efficiency, as compared to topologies presented in [6] and [7] depicted in figure 12. Figure 11 represents the comparison between transformer's secondary voltages for all three topologies The fast recovery SIC diodes present at the secondary side of the converter, gives the proposed converter a sinusoidal-like waveform at the transformer secondary as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:55:16 UTC from IEEE Xplore. Restrictions apply.

compared to topologies presented in [6] and [7] shown in figure 11.

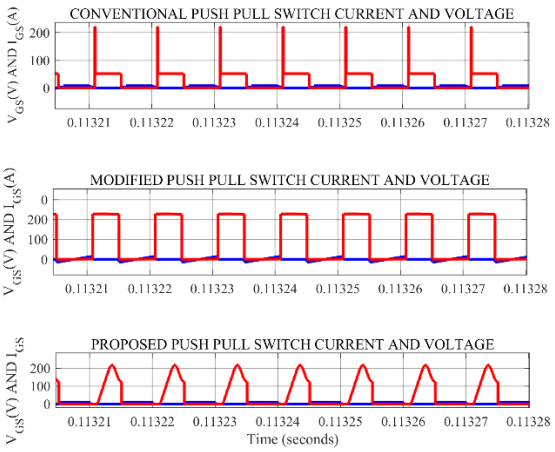


Fig. 10 Comparative waveform of drain to source voltage and current for three topologies

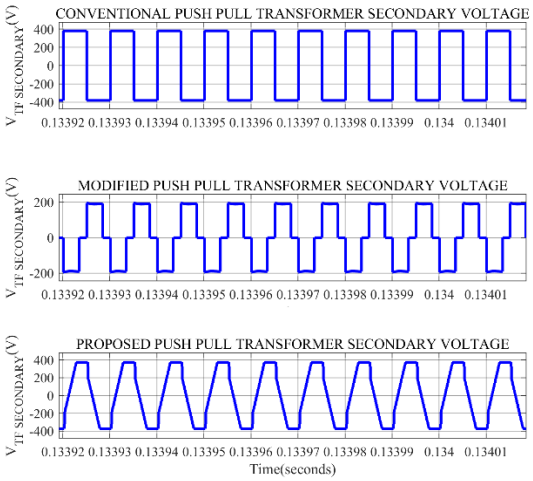


Fig. 11 Comparative waveform of transformer secondary voltage for three topologies

VI. Conclusion

The active clamped ZCS-ZVS resonant push–pull converter has been proposed. The analysis, design and simulation results for the proposed converter have been presented. High voltage conversion ratio, reduced switch conduction loss, and other features of the typical current-fed push-pull converter are all preserved in the proposed converter. Moreover, the proposed converter outperforms all other examined topologies with an efficiency of 97.135. In particular, the ripple current of the boost inductor is extremely low at the duty ratio around 0.58. The proposed converter is suitable for the high step-up converter in fuel cells and solar cell applications due to features like soft switching, sinusoidal shape voltage waveform at the transformer primary, and sinusoidal secondary voltage of the transformer due to fast recovery SIC diode, and single gate requirement.

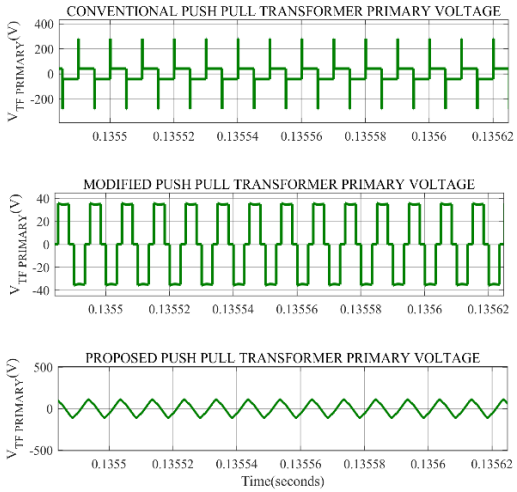


Fig. 12 Comparative waveform of transformer primary voltage for three topologies

References

[1] Q. Wu, M. Wang, W. Zhou, C. Huang, G. Liu and X. Wang, "High Frequency Active-Clamped Zero-Current Switching Current-Fed Push-Pull Converter for Micro-Converter Applications," 2020 IEEE Energy Conversion Congress and Exposition (ECCE), Detroit, MI, USA, 2020, pp. 1273-1278, doi: 10.1109/ECCE44975.2020.9236043.

[2] K. R. Sree and A. K. Rathore, "Impulse Commutated Zero-Current Switching Current-Fed Push–Pull Converter: Analysis, Design, and Experimental Results," in IEEE Transactions on Industrial Electronics, vol. 62, no. 1, pp. 363-370, Jan. 2015, doi: 10.1109/TIE.2014.2331029.

[3] Qunfang Wu, Qin Wang, Jialin Xu, Zilong Xu, "Active-clamped ZVS current-fed push–pull isolated dc/dc converter for renewable energy conversion applications"in IET Power Electronics,Nov. 2017, doi: 10.1049/iet-pel.2017.0144.

[4] P. Xuewei and A. K. Rathore, "Naturally Clamped Zero-Current Commutated Soft-Switching Current-Fed Push–Pull DC/DC Converter: Analysis, Design, and Experimental Results," in IEEE Transactions on Power Electronics, vol. 30, no. 3, pp. 1318-1327, March 2015, doi: 10.1109/TPEL.2014.2315834.

[5] L. Jiang et al., "A New Push-Pull DC/DC Converter Topology With Complementary Active Clamped," in IEEE Transactions on Industrial Electronics, vol. 69, no. 6, pp. 6445-6449, June 2022, doi: 10.1109/TIE.2021.3088357.

[6] QUN Z., LEE F.C.: 'High-efficiency, high step-up DC–DC converters', IEEE Trans. Power Electron., 2003, 18,pp. 65– 73.

[7] E.-H. Kim, B.-H. Kwon, "High step-up resonant push–pull converter with high efficiency"in IET Power Electronics, June 2007, doi: 10.1049/iet-pel:20070140.

[8] O. Abdel-Rahim, A. Chub, A. Blinov and D. Vinnikov, "Current-Fed Dual Inductor Push-Pull Partial Power Converter," 2022 IEEE 20th International Power Electronics and Motion Control Conference (PEMC), Brasov, Romania, 2022, pp. 327-332, doi:10.1109/PEMC51159.2022.9962937.

[9] B. A. Basit, M. Kim and J. -W. Jung, "Improved Duty Compensation Control-Based Bidirectional Resonant DC–DC Converter with Reduced Input-Current Ripple for Battery Energy Storage Systems," in IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2023.3244349.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:55:16 UTC from IEEE Xplore. Restrictions apply.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Compact BERT-Based Multi-Models for Efficient Fake News Detection

Vineet Rana
Dept. of Computer Engineering
Delhi Technological University
Delhi, India
vineetrana_2k19co430@dtu.ac.in

Vyom Garg
Dept. of Computer Engineering
Delhi Technological University
Delhi, India
vyomgarg_2k19co438@dtu.ac.in

Yatharth Mago
Dept. of Computer Engineering
Delhi Technological University
Delhi, India
yatharthmago_2k19co446@dtu.ac.in

Dr. Aruna Bhat
Dept. of Computer Engineering
Delhi Technological University
Delhi, India
aruna.bhat@dtu.ac.in

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205773

***Abstract*—In today's world, online news has become an integral and important part of the daily lives of millions of beings. With every day gone by, an increasing number of people have begun to rely on various websites for their daily dose of information. Due to the ever increasing popularity of these sources, the scope of spreading misinformation has also risen to a great extent. Rumors and biased news can spread like wildfire, thus undermining the credibility of news articles and causing trust in the media to dwindle. This recent surge in false information on social media has prompted extensive research in the field of online fake news detection. Hence, there is a need for efficient yet accurate fake news detection models which can be made accessible to everyone. In this research, we have explored the use of pre-trained distilled variants of the BERT model to identify fake news and rumors from actual online news. We compared the results before and after combining two smaller pre-trained BERT models into a single architecture with other approaches to see how well distilled models are able to learn the key features required to distinguish fake news from real. Results show that these smaller variants were comparable to the related work in terms of accuracy while still being efficient, compact, and easier to train.**

***Index Terms*—ake News Detection, Deep Learning, Natural Language Processing, BERT, Pre-trained Language Modelake News Detection, Deep Learning, Natural Language Processing, BERT, Pre-trained Language ModelF**

I. Introduction

Fake news and the media's stance have always been a matter of concern for societies, as they have the potential to incite political unrest and could generate panic among the masses. The COVID-19 pandemic has only exacerbated the problem, as there has been an overabundance of misinformation and fake news about the virus circulating on social media platforms. With the advent of social media, fake news has become more rampant and has the potential to spread at an alarming rate, leading to what is now known as an "infodemic."

To manage the spread of such misinformation, we need tools that can derive linguistic, sentimental, emotional, and other characteristics from different news articles. In this regard, transformer-based bi-directional models like BERT have shown impressive results. However, the computational requirements of BERT are quite high, making it difficult to use in applications that require real-time processing. In order to address this issue, smaller variants of BERT have been developed, such as BERT-small and BERT-tiny. These models have been shown to be computationally efficient while still maintaining a high level of accuracy. BERT forms the basis of this research, along with its computationally efficient counterparts such as BERT-small and BERT-tiny, both of which will be explored throughout the course of the study, along with their multi-model variants, in order to analyze online fake news, by applying transfer learning.

II. Related Works

Numerous approaches have been explored to handle the task of online news analysis, ranging from the detection of fake news to news sentiment analysis.

The authors of [1] proposed a multi-task model for fake news detection named FDML. This was based primarily on two observations i.e., a few topics may have a higher percentage of fake news, and some writers may have greater intentions to publish fake news. The FDML was claimed to be one of the first works that integrated the task of fake news detection along with classifying the news topics too.

In [2], Julio CS et al. experimented with a wide range of news article attributes such as textual features, news source features, environmental features, etc., and then further used the same with different classifiers, with XGBoost & Random Forest showing superlative performance.

Kai et al. in [3] focused on a broad understanding of fake news detection and its characterizations based on social theories and psychology, along with analyzing existing algorithms from data mining's viewpoint.

In [4], the authors made use of a model known as FNDNet, which makes use of a deep neural network's several hidden layers in order to learn the distinctive attributes of classification of fake news. This proposed model achieved a whopping

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:03 UTC from IEEE Xplore. Restrictions apply.

accuracy of over 98% on the Kaggle fake news dataset, and the results were validated using various performance parameters.

The authors of [5] made use of the Naive Bayes Classifier so that they could differentiate between fake and real news. This describes in detail how the process of spam filtering goes about and the features associated with it. On the Buzzfeed news dataset, the model displayed an accuracy of 75%.

III. Methodology

A. Transformers

A major chunk of the learning systems up until recently processed natural language using gated RNNs (for example, LSTM & GRU) which are made for processing sequential input. These designs made an important addition of the attention layers to their architecture after inculcating the attention mechanism for better understanding the relationships between the sequence's components and then weighing them in accordance with their significance to the model [8, 9]. Transformers have mostly replaced the prior sequence models in recent times because of their superior performance in tasks related to Natural Language Processing. The attention mechanism is capable of satisfying the RNN's predicted learning performance, so transformers make use of the same methodology [10]. One of the most promising and latest transformer models is the BERT, which is pre-trained on a large amount of unlabeled data for two purposes [11]:

- Masked language modeling: In a particular sentence or phrase, around 15% of the words are hidden (or masked) so that the model can forecast them based on word context. It should be emphasized that BERT is a bidirectional model, which means that it uses both pre-word and post-word content to predict, in contrast to earlier Transformer models like the GPT. Additionally, both left-to-right and right-to-left languages can use BERT with equal effectiveness, which signifies that it does not rely on the direction of the language.
- Next sentence prediction: As the name suggests, whenever a pair of sentences is given, the model should be capable of predicting if the second sentence follows the first one or not.

The pre-trained model needs fine-tuning in order to be used in downstream activities. The same can be done by applying it to task-specific data, keeping the architecture and weight same as before. Only after the task's intended output is met do we add the output layers. We initially start with the model parameters from the pre-training part when fine-tuning the model. All model parameters are then adjusted simultaneously.

B. Transfer Learning

Deep learning needs an abundance of data to learn and train from, but our access to rumors and false news is quite constrained. This presents a fundamental problem for employing deep neural networks on fake news, because of the limited amount of resources. The majority of the fake news datasets that have been developed to date are either a few hundred or a few thousand rows in size, which makes them unsuitable for training complicated neural networks with many parameters [12, 13]. Transfer learning has been employed in recent studies to solve this issue. In transfer learning, a model, called generally as the pre-trained model, is first trained on a sizable quantity of generic data, usually extracted from the internet. Then, using labeled data for a particular task that shares some knowledge with the core model, the pre-trained model is enhanced [14]. This kind of learning is recently being inculcated in NLP, after being used for several years in image processing methods.

Typically, the pre-training procedure is used to create a language model for transfer learning in Natural Language Processing. A Language models' main task is to anticipate the text's next word or phrase. Language models contain two key characteristics: Firstly, they can be taught on unlabeled data, giving them access to a tonne of data for a thorough learning process. Secondly, a language model which is well-trained may provide us with a thorough knowledge of the semantics and structure of the language, later which becomes crucial for a number of tasks.

C. Distillation

Due to their enormous size, training large-sized models like BERT might take much more time, ranging from a few hours to even days. Additionally, you can discover that running them on small devices would result in significant memory and time expenditures due to today's constantly improving performances.

Distillation can be summarized as an act of teaching a tiny student model to closely resemble a bigger instructor model. Distillation is pretty useful when we intend to migrate a model to less powerful hardware, like a smartphone or a tiny laptop, because distilled models are more efficient and take up less space.

D. Smaller BERT Variants

Two smaller versions of BERT, called BERT-Tiny and BERT-Small, have been developed to address the issue of limited computational resources. These models are easier and faster to train than the original BERT model, as they have fewer parameters as compared to the 110M parameters present in the original model, while still maintaining a high level of accuracy. BERT-Tiny is a miniature version of the original BERT model, with only 2 layers and 128 hidden units per layer. BERT-Tiny is suitable for small-scale NLP tasks that do not require a large amount of computational power.

BERT-Small is a slightly larger version of BERT-Tiny, with 4 layers and 512 hidden units per layer, making it a good choice for medium-scale NLP tasks. BERT-Small is also faster and more efficient than the original BERT model, while still achieving comparable accuracy. BERT-Tiny and BERT-Small, together with the other variations, were proposed in "Well-Read Students Learn Better: On the Importance of Pre-training Compact Models", and then were ported to "Generalization in NLI: Ways (Not) To Go Beyond Simple Heuristics" [18,19].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:03 UTC from IEEE Xplore. Restrictions apply.

We used the BERT-Tiny and BERT-Small models from the Hugging Face transformer library [22].

E. Dual Model Architecture

The main idea behind the proposed dual model architecture is to train two BERT variants jointly on a single task, with each model having different parameters and being initialized randomly. This approach not only allows for the use of smaller, more efficient models but also leverages the complementary strengths of the two models to improve classification accuracy.

We have utilized Dual BERT-Tiny and Dual BERT-Small, which are specifically designed to be computationally efficient and require less memory compared to larger models, while still delivering good performance on various natural language processing tasks. Inputs are encoded as 2 tuples of matrices (token, mask, sequence ids), one for each input. The first model receives the news headlines while the second one receives the body of the news article or a brief textual description. The output layers of both transformers were then concatenated and passed through a fully connected layer to obtain the final output, which was used for fake news classification.

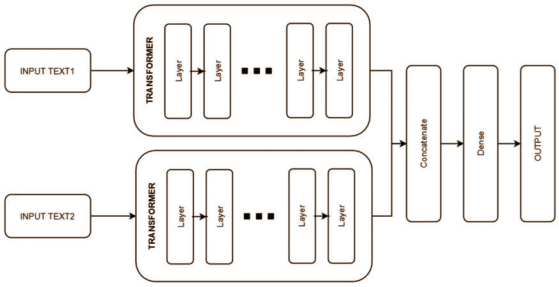


Fig. 1. Architecture of Dual BERT variant

IV. Experiments and Evaluation

A. Dataset

FNC-1 was introduced as a dataset for stance detection and it consists of labeled pairs of a headline text and a body text, which can be either from the same news article or from two different articles. These pairs are labeled as agree (the body text and the headline are consistent), disagree (the body text and the headline are not consistent), discuss (the body text discusses the same subject as the headline, but it is neutral), and unrelated (the headline discussed is different than the body content). The dataset's headlines are all written in the form of statements and contain 25413 samples in the test set and 49972 in the training set.

B. Evaluation Metrics

To evaluate the performance of our models, we consider some of the widely used metrics: accuracy, precision, recall, and F1 score. The number of correctly predicted positive classes by the model is denoted by True Positive (TP), while the number of correctly predicted negative classes is represented by True Negative (TN). False Positive (FP) represents the number of positive classes classified as false by the model, and False Negative (FN) refers to the number of negative classes classified as true by the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

C. Training parameters

To address the class imbalance in the training dataset, we used focal loss as the loss function and assigned more weights to classes such as *agree*, *disagree*, and *discuss* based on their distribution [21]. During the initial phase of training, the pre-trained layers in Dual BERT-Tiny and Dual BERT-Small models were locked, and then unlocked once the validation loss started to stabilize. This approach allowed the models to better adapt to the specific task of fake news classification while preventing the integrity of the pre-trained weights.

D. Results

We compared our result with the top 4 submissions of teams who participated in the FNC-1 challenge, which are shown in the first half of Table-1, resulting in Athene [28] having the highest accuracy. Per class accuracy is shown in columns 3-6 in % while the second column shows the overall accuracy. The BERT-Tiny and BERT-Small were able to give an accuracy of 85.25%, and 90.11% respectively, and the Dual BERT-Tiny and Dual BERT-Small gave an accuracy of 82.79% and 79.96% respectively on the test set. Table-2 shows precision, recall, and F1-score in % for each class obtained using the BERT-Small model.

TABLE I
Results on FNC-1 Test Dataset

Method	Accuracy	Per Class Accuracy			
		Unrelated	Discuss	Agree	Disagree
SOLAT in the SWEN [27]	89.08	98.70	76.18	58.50	1.86
Athene [28]	89.50	**99.25**	80.89	44.72	9.47
UCL Machine Reading [29]	88.46	97.90	**81.38**	44.04	6.60
Chips Ahoy! [30]	88.01	98.98	70.29	55.96	0.28
BERT-Tiny	85.25	96.19	69.65	45.14	6.89
Dual BERT-Tiny	82.79	93.28	69.85	41.25	3.01
BERT-Small	**90.11**	98.95	76.28	**65.53**	**12.91**
Dual BERT-Small	79.96	92.14	59.21	38.62	5.16

Conclusion and Future Works

In this research, we first addressed the need for an efficient mechanism to differentiate online fake news from real, authentic news. We then explored the use of deep transformer-based bi-directional models for the analysis of online fake news, which showed impressive results across the FNC-1 dataset.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:03 UTC from IEEE Xplore. Restrictions apply.

TABLE II
BERT-SMALL RESULTS ON FNC-1 TEST DATASET

Stance	Precision	Recall	F1-score
unrelated	97.98	98.95	98.47
discuss	80.00	76.28	78.10
agree	52.66	65.53	58.39
disagree	34.88	12.91	18.85

Our findings suggest that comparatively smaller pre-trained models such as BERT-Tiny and BERT-Small, when coupled with appropriate loss functions, can deliver great results when used for fake news detection while also being more efficient and faster to train.

Moving forward, further exploration of linguistic, sentimental, emotional, and other characteristics of online news could lead to even better results when combined with fake and real news labels. By analyzing these additional characteristics, we can improve our understanding of the nature of fake news and develop more accurate and effective detection methods. Additionally, the effectiveness of the proposed method could be tested on other datasets and in different languages to evaluate its generalizability. Overall, the findings of this study have important implications for the detection and prevention of fake news, which is a critical issue in today's digital world.

REFERENCES

[1] Liao, Qing, et al. "An integrated multi-task model for fake news detection." IEEE Transactions on Knowledge and Data Engineering (2021).

[2] Reis, Julio CS, et al. "Supervised learning for fake news detection." IEEE Intelligent Systems 34.2 (2019): 76-81.

[3] Shu, Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter 19.1 (2017): 22-36.

[4] Kaliyar, Rohit Kumar, et al. "FNDNet–a deep convolutional neural network for fake news detection." Cognitive Systems Research 61 (2020): 32-44.

[5] Granik, Mykhailo, and Volodymyr Mesyura. "Fake news detection using naive Bayes classifier." 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON). IEEE, 2017.

[6] Ahmed H, Traore I, Saad S. "Detecting opinion spams and fake news using text classification", Journal of Security and Privacy, Volume 1, Issue 1, Wiley, January/February 2018.

[7] Ahmed H, Traore I, Saad S. (2017) "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques. In: Traore I., Woungang I., Awad A. (eds) Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham (pp. 127-138).

[8] D. Bahdanau, K.H. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 2015.

[9] M.T. Luong, H. Pham, C.D. Manning, Effective approaches to attention-based neural machine translation, in: Conf. Proc. - EMNLP 2015 Conf. Empir. Methods Nat. Lang. Process., 2015. https://doi.org/10.18653/v1/d15-1166.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Adv. Neural Inf. Process. Syst., 2017

[11] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf. 1 (2019) 4171– 4186.

[12] F. Pierri, S. Ceri, False news on social media: A data-driven survey, SIGMOD Rec. 48 (2019) 18–32. https://doi.org/10.1145/3377330.3377334.

[13] W.Y. Wang, "Liar, liar pants on fire": A new benchmark dataset for fake news detection, in: ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap., 2017. https://doi.org/10.18653/v1/P17-2067.

[14] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2009) 1345–1359.

[15] Sanh, Victor & Debut, Lysandre & Chaumond, Julien & Wolf, Thomas. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.

[16] Tenney I, Das D, Pavlick E (2019) BERT rediscovers the classical NLP pipeline. In: Proceedings of the 57th annual meeting of the association for computational linguistics

[17] Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018) Deep contextualized word representations. In: Proceedings of NAACL-HLT, pp 2227–2237

[18] Turc, Iulia, et al. "Well-Read Students Learn Better: On the Importance of Pre-training Compact Models." arXiv.org, 23 Aug. 2019, arxiv.org/abs/1908.08962v2.

[19] Bhargava, Prajjwal, et al. "Generalization in NLI: Ways (Not) to Go Beyond Simple Heuristics." arXiv.org, 4 Oct. 2021, arxiv.org/abs/2110.01518v1.

[20] A. Bhat, R. Mahar, R. Punia and R. Srivastava, "Exploring Multimodal Sentiment Analysis through Cartesian Product approach using BERT Embeddings and ResNet-50 encodings and comparing performance with pre-existing models," 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022, pp. 1-6, doi: 10.1109/INCET54531.2022.9825245.

[21] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980–2988).

[22] Wolf, T. et al. (2019). "HuggingFace's Transformers: State-of-the-art Natural Language Processing". In: ArXiv abs/1910.03771.

[23] A. Bhat, V. Varshney, V. Bajlotra and V. Gupta, "Detection of hatefulness in Memes using unimodal and multimodal techniques," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 65-73, doi: 10.1109/ICICCS53718.2022.9788376.

[24] Monika and A. Bhat, "An analysis of Crime data under Apache Pig on Big Data," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2019, pp. 330-335, doi: 10.1109/I-SMAC47947.2019.9032565.

[25] Isaac, Akileng, and Aruna Bhat. "A Conceptual Enhancement of LSTM Using Knowledge Distillation for Hate Speech Detection." In Smart Trends in Computing and Communications: Proceedings of SmartCom 2021, pp. 561-570. Springer Singapore, 2022.

[26] M. Ahmed, M. Goel, R. Kumar and A. Bhat, "Sentiment Analysis on Twitter using Ordinal Regression," 2021 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Pune, India, 2021, pp. 1-4, doi: 10.1109/SMARTGENCON51891.2021.9645751.

[27] S. B. Yuxi Pan, Doug Sibley. Talos. http://blog.talosintelligence.com/2017/06/, 2017.

[28] B. S. Andreas Hanselowski, Avinesh PVS and F. Caspelherr. Team athene on the fake news challenge. 2017

[29] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel. A simple but tough-to-beat baseline for the fake news challenge stance detection task. arXiv preprint arXiv:1707.03264, 2017.

[30] J. Shang. Chips ahoy! at fake news challenge. 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:03 UTC from IEEE Xplore. Restrictions apply.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Comparative Analysis Of Machine Learning Models For Fake News Classification

Archit Gupta
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
architgupta1202@gmail.com

Arnav Batla
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
batlaarnav@gmail.com

Chaitanya Kumar
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
kumarchaitanya1108@gmail.com

Dr. Goonjan Jain
Department of Applied Mathematics
Delhi Technological University
New Delhi, India
goonjanjain@dtu.ac.in

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205870

***Abstract*—It has become clear that fake news is dangerous. Identifying fake news is a crucial step towards preserving the virtue and prosperity of society. Social media's rising popularity has led to an increase in the spread of false information. There aren't enough frameworks in place to deal with misleading news. There are many low-cost online news sources and it's an easy access via social media. These are the reasons why there's a spread of fake news. News Content is the only reason for the present fake news detection algorithms, also users' previous posts or activities provide a lot of insights about their views on news and have a significant effect on false news identification. The proposed research seeks to investigate several machine learning approaches for the analysis and identification of false news. In order to identify the spread of fake news on social media, we compare various widely used machine learning methods, such as Naive Bayes and Multi Layer Perceptron Classifiers, in this study. In this work, using solely text data, we develop a number of machine learning methods using the WELFake dataset.**

***Index Terms*—fake news detection, machine learning, text analysis, naive bayes, multi layer perceptron**

I. Introduction

Nowadays, social media networks are the most popular way to distribute thousands of pieces of news, whether they are public or private. Content on social media is simple to obtain, share, and remark on. There is a danger of getting exposed to false information which is purposefully wrong or contains inaccurate information in order to support a certain political or economic objective as it is simpler for users or readers to share their particular opinions. Furthermore, false information typically spreads more quickly, deeply, and widely on social media.

In the past year, fake news detection in social media has drawn a lot of attention from academic and professional circles. Social media networks have identified numerous web pages and misinformation and have devoted resources to the job. A fake news detection method aims to identify and investigate various types of potentially misleading news. The analysis of prior real and fake news samples is used to predict the likelihood that a given piece of news content is being intentionally misleading.

A notable obstacle to this algorithmic method and resolving this natural language processing issue is the lack of news examples that are available as corpora for training the model. The various false news models that already existed were context-specific. There is a dearth of a suitable paradigm for classifying the different kinds of deceptions that can happen when working with written data. This study looks at various techniques and kinds of deception that can be found when working with online news material, weighing the benefits and drawbacks of predictive modeling. It offers an algorithmic method for resolving the specified issue.

We used the WELFake dataset, which consists of 72,134 news stories, 35,028 of which are true and 37,106 of which are false. In order to avoid over-fitting of classifiers and to give more text data for better ML training, authors combined four well-known news datasets (Kaggle, McIntire, Reuters, and BuzzFeed Politics). The four columns in this dataset are Serial number (beginning at 0), Title (about the text news headline), Text (about the news substance), and Label (0 = fake, 1 = authentic). Just 72134 of the 78098 data elements in the csv file are accessible according to the data frame.

The primary goal of this study is to minimise and decrease the spread of false information on social media platforms. It is more practical for the consumer when the virtual site is secure. In this research, we'll attempt to put some methods into practise to see how well the algorithms performs at detecting authenticity. We think it will aid in the identification of rumours and have a transformative effect on the veracity of social media news. Additionally, it will stop people from trusting anything they read or see on social media.

II. Related Works

Akshay Jain and Amey Kasbe conducted a research study in which they evaluated the accuracy of various machine learning algorithms for making predictions. They tested models such as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:33 UTC from IEEE Xplore. Restrictions apply.

bounded decision trees, gradient boosting, and support vector machines, using an unreliable probability threshold. The study found that these models achieved an accuracy ranging between 0.85 to 0.91.[1]

In their study, Farzana Islam and her colleagues utilized the Naïve Bayes classifier algorithm method to categorize deceptive news, employing both count vectorization and TF-IDF vectorization as feature extraction methods. They applied this technique to two open source fake news datasets available on various sources[2].

Zongru Shao and Pranav Bharadwaj conducted experiments on the fake news dataset available on Kaggle.com, utilizing RNN, , a random forest classifiers and a Naïve Bayes classifier. They employed various feature extraction techniques, including Quadgram, Bigram, Trigram, and GloVe[3].

Castillo et al. used a different and more logical approach in their research. They identified positive and negative keywords and hashtags to identify the legitimacy of the news[4].Karimi and Tang introduced a novel technique to detect fabricated news, which relies on HDSF more commonly known as Hierarchical Discourse-level Structure. It is a hierarchical structure. In this algorithm, features of both authentic and fake news documents are pulled out automatically. The researchers evaluated this approach on five datasets [5].

Z Khanam et al used feature selection and processing of attributes to apply machine learning models like SVM, Logistic regression, KNN, XGBoost and obtained the accuracy of 0.92 using SVM model on the dataset imported from Kaggle[6].

Veronica Perez-Rosas used different models, including the decision tree, naive Bayes algorithms and clustering were employed to differentiate between Twitter spam senders. On average, the models were able to identify spammers with an accuracy of 0.70 and fraudsters with an accuracy of 0.712. However, the models exhibited a low level of intermediate precision in distinguishing spammers from non-spam accounts[7].

Their study assumes that Fake news detection is an application of predictive analysis that involves three stages: cleaning/processing, feature extraction, and classification or categorising. A hybrid classification model designed specifically for identifying counterfeit news was used in this study. The classification approach involved a combination of random forests and KNN. Metrics like accuracy, precision and recall for the proposed models were calculated, and the results showed an improvement of up to 8 percent using a hybrid model for fake news detection[8].

Alim Al Ayub Ahmed and team in their research tried to identify which models can and can not be used for fake news classification and the steps to implement them[9]. Fahim Belal Mahmud and team applied models like SVM, LR, Random forest using the spacey and BERT techniques and using four GNN vaiants : GAT,GraphSAGE, GCN, and GIN.[10]

III. METHODOLOGY

This study utilized a three-stage approach for its methods. In the first stage, the raw dataset underwent pre-processing which involved filtering and various data cleaning techniques to extract semantic features. A stopword filter was implemented to remove prepositions and categorize the data. The second stage involved the conversion of text based features into binary vectors through numerical techniques. The final stage utilized both machine learning and deep learning algorithms like Naive Bayes, Random Forest, XGB Classifier to create partitions within the dataset. The same methodology is shown through figure 1.

A. The Pre-processing Stage

Data preprocessing involves the cleaning and transformation of unstructured data, such as text, to prepare it for analysis. Text mining is a common approach to this task, but it can be challenging due to the presence of impurities like HTML tags, single characters, numbers, apostrophes and unwanted symbols. In this study, we applied the stopwords removal method to clean and preprocess a classified dataset. Stopword removal is a widely used data filtering technique in information retrieval and text classification. It involves eliminating certain common words like: ”the, in, a, an, with”, that have little impact on the categorization of data.

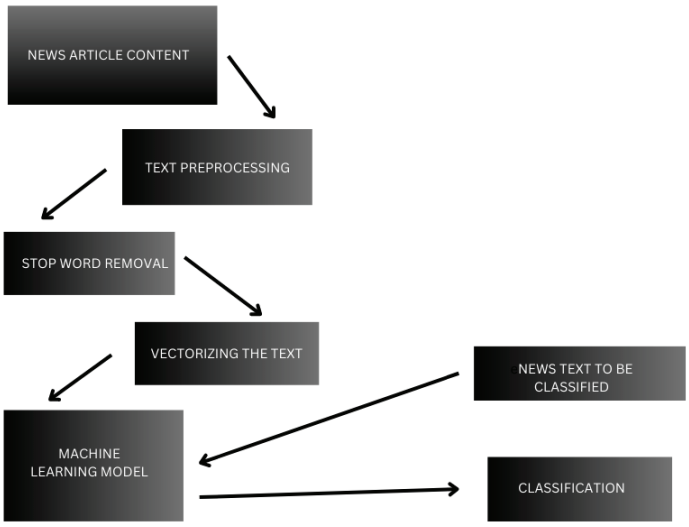


Fig. 1. Structure of this Paper

B. The Extract Features Stage

In this technique we use a method to convert the textual data to binary vectors 0 and 1 to make it convenient to train the classification algorithms. New vectors are created from each new entry in the sample text file. Vectors can be created using multiple techniques mentioned below:

1. TF-IDF vectorizer: The term frequency-inverse document frequency (TF-IDF) vectorizer is one of the most common method for the first stage i.e Feature Extraction. This technique involves two stages, with the first stage calculating the term frequency (TF), and the second stage computing the inverse document frequency (IDF).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:33 UTC from IEEE Xplore. Restrictions apply.

Term Frequency = (Number of times the term appears in a document) / (Total number of terms in the document)

Inverse Document Frequency = log(Total number of documents / Number of documents with the term)

2. N-gram level vectorizer: An n-gram level vectorizer is a technique used in natural language processing (NLP) to convert text data into numerical vectors that can be used for machine learning. N-grams are contiguous sequences of n words from a given text, and the vectorizer creates a feature matrix with the frequency count of each n-gram in the text. This approach captures the contextual information of the text, making it useful for various NLP tasks like sentiment analysis and text classification. N-gram level vectorizers are customizable and can generate feature matrices with different levels of granularity.

3. Character level vectorizer: A character-level vectorizer is a technique used in natural language processing to convert text data into numerical vectors by considering each character in the text as a separate feature. This approach captures the underlying structure and patterns of the text at a granular level and is useful for tasks such as language modeling and text generation.

C. The Classifiers Stage

The following Machine learning and deep learning algorithms were used in this study: Gaussian Naive Bayes, Bernoulli Naive Bayes, Multilayer Perceptron, Random Forest Classifier, XGBoost Classifier.

- **Gaussian Naive Bayes** is a classification algorithm that use Bayes' theorem and assumes that features are independent and normally distributed. It estimates the mean and variance of the input features for each class to compute the probability of each class given the input features. Using these estimations, the system then computes the likelihood of each feature value given the class. It derives the joint probability of the input feature vector given the class by multiplying the likelihoods for each feature. Finally, given the input feature vector, it computes the posterior probability of each class and selects the class with the highest probability as the predicted class.
- **Bernoulli Naive Bayes** is based on the assumption that the features are binary or boolean values. It estimates the likelihood of each feature for each class to compute the probability of each class given the supplied features. Using these probabilities, the algorithm then computes the likelihood of the input feature vector given the class. It derives the joint probability of the input feature vector given the class by multiplying the likelihoods for each feature. Finally, given the input feature vector, it computes the posterior probability of each class and selects the class with the highest probability as the predicted class. Bernoulli Naive Bayes is a text classification method in which the features indicate the presence or absence of specific words in a document.
- **MLP (Multilayer Perceptron) Classifier** is an artificial neural network that is commonly utilized for supervised learning tasks such as classification. It comprises of several layers of interconnected nodes, where each node carries out a weighted sum of its inputs and applies an activation function to generate its output. MLPs are extensively used in various applications such as image recognition, speech recognition, and natural language processing. They have the ability to learn intricate non-linear connections between inputs and outputs and can be trained using the backpropagation algorithm.
- **Random Forest Classifier** is an ensemble learning technique that enhances the accuracy and consistency of predictions by combining multiple decision trees. It creates numerous decision trees during training and then determines the class by identifying the mode of the predicted classes by individual trees. Random Forests have the capability to handle large datasets with high dimensionality and learn non-linear relationships between input features and output classes. They are commonly used in fields such as finance, medicine, and e-commerce for various applications like fraud detection, disease diagnosis, and customer segmentation. Moreover, Random Forests provide feature importance scores that can be used for data interpretation and feature selection.
- **XGBoost (Extreme Gradient Boosting) Classifier** algorithm is a robust machine learning technique that is extensively used for supervised learning problems like classification and regression. It is a highly optimized implementation of the gradient boosting algorithm, which uses a combination of multiple decision trees to enhance the accuracy and efficiency of predictions. XGBoost applies a gradient descent algorithm to optimize the objective function and prune the decision trees to avoid overfitting. This algorithm can handle missing values and feature interactions, and it supports parallel processing to expedite model training. XGBoost is utilized in various fields such as finance, healthcare, and marketing for tasks like credit risk assessment, disease diagnosis, and customer churn prediction.

D. Results

The primary objective of this project is to perform a comprehensive analysis of the news data contained in the WELFake dataset, which is available on the Kaggle website. Our analysis involves the use of six distinct algorithms, namely:

- Bernoulli Naive Bayes
- MLPClassifier
- Gaussian Naive Bayes
- Random Forest Classifier
- XGBoost

These algorithms are chosen due to their proven effectiveness in the data driven field like data analysis and machine learning.

We conduct our analysis by first importing the dataset into the Anaconda platform and then executing the algorithm codes in Python. The cognitive learning library is used

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:33 UTC from IEEE Xplore. Restrictions apply.

to automatically generate a confusion matrix, which we use to visualize the results of our analysis. Additionally, performace was measured based on the calculated value of accuracy of each algorithm.

The confusion matrix and accuracy scores for each of the six algorithms used in the analysis are presented below. As can be seen, the XGBoost algorithm outperforms all the others with a high accuracy score of 94%, while the Gaussian Naive Bayes algorithm has the lowest accuracy score of 74%. The results obtained from our analysis provide valuable insights into the usefulness of different algorithms for detecting fake and deceptive news and can be used to inform future research in this area.

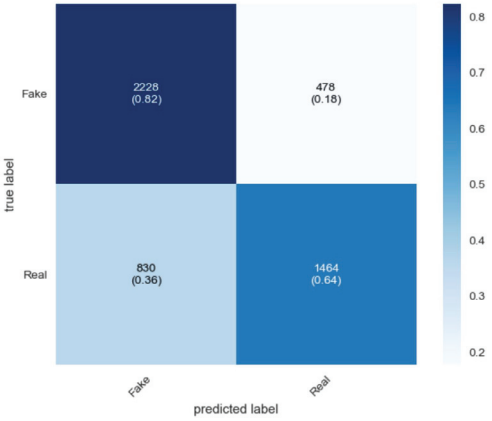


Fig. 2. Confusion Matrix for Gaussian Naive Bayes.

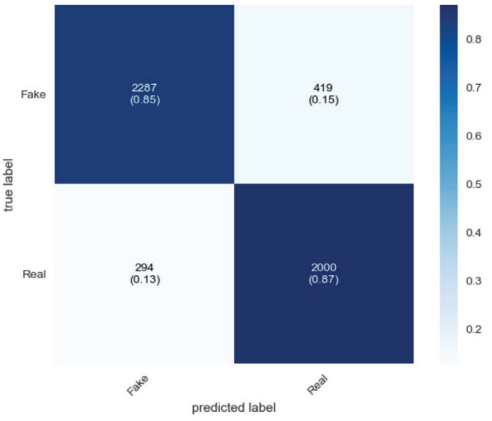


Fig. 3. Confusion Matrix for Bernoulli Naive Bayes.

CONCLUSION

The primary focus of this research paper is to address the growing issue of fake news on the internet, specifically

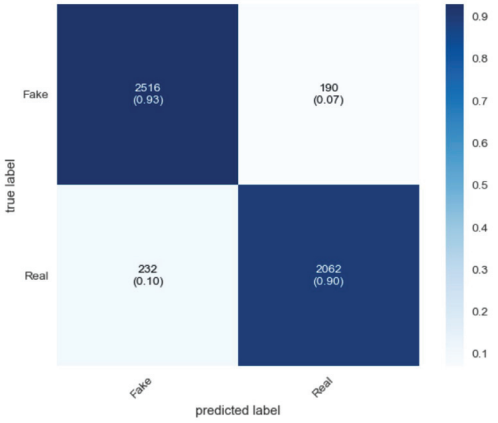


Fig. 4. Confusion Matrix for Multi Layer Perceptron.

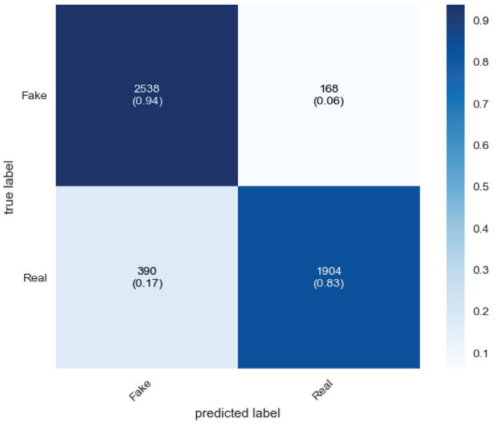


Fig. 5. Confusion Matrix for Random Forest Classifier.

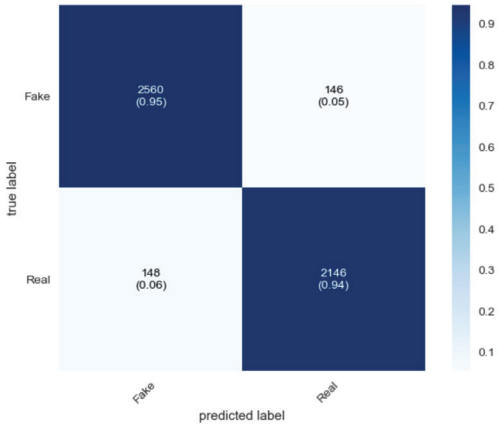


Fig. 6. Confusion Matrix for XG Boost Classifier.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:33 UTC from IEEE Xplore. Restrictions apply.

TABLE I
RESULTS

Model	*Accuracy(Train)*	*Accuracy(Test)*	*CV Score Mean(Train)*
GaussianNB	1.00	0.74	0.72
BernoulliNB	0.91	0.86	0.85
MLPClassifier	1.00	0.92	0.90
RandomForest	1.00	0.89	0.89
XGBClassifier	1.00	0.94	0.95

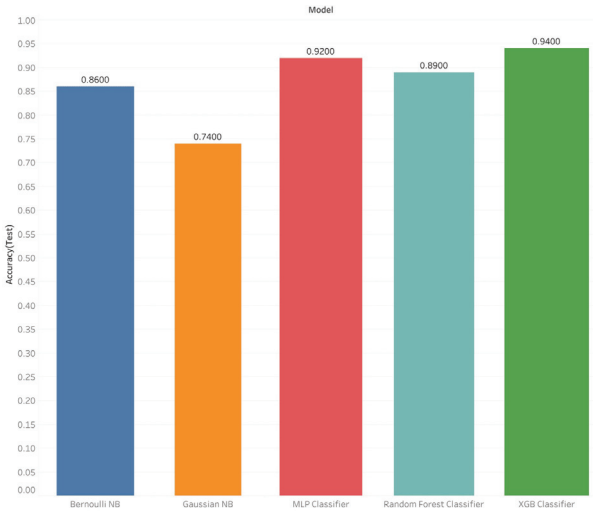


Fig. 7. Accuracy Scores

regarding textual content. The paper aims to provide a comprehensive analysis of previous studies and advancements in the field, and to conduct a comparative analysis on the WELFake dataset available on Kaggle.com. The main objective of the study is to determine the accuracy of various algorithms in classifying news as either real or fake. To achieve this, five different algorithms were applied to the dataset: GaussianNB, BernoulliNB, MLPClassifier, RandomForestClassifier, and XGBClassifier.

The findings of the study indicate that XGBClassifier exhibited the highest accuracy, while GaussianNB demonstrated the lowest accuracy among the tested algorithms. However, the study's results are limited due to the use of only one dataset, and there is significant potential for further research and development of new algorithms to expand the classification beyond textual content to non-textual media such as fake images or videos.

In conclusion, the study underscores the importance of developing effective algorithms to combat the proliferation of fake news on the internet. The use of advanced algorithms in conjunction with human oversight can significantly improve the accuracy and reliability of news classification, thereby reducing the impact of fake news on society.

REFERENCES

[1] Akshay Jain and AmeyKasbe. "Fake News Detection." 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS). Bhopal, India: IEEE.2018.

[2] A. Farzana Islam, et al. "Effect of Corpora on Classification of Fake News using Naive Bayes Classifier." International Journal of Automation, Artificial Intelligence and Machine Learning 1.1 (2020): 80-92.

[3] P. Bharadwaj, and Z. Shao. "Fake news detection with semantic features and text mining." International Journal on Natural Language Computing (IJNLC) Vol 8 (2019).

[4] C. Carlos, M. Mendoza, and Barbara Poblete. "Information credibility on twitter." Proceedings of the 20th international conference on World wide web. 2011.

[5] K., Hamid, and Jiliang Tang. "Learning hierarchical discourse-level structure for fake news detection." arXiv preprint arXiv:1903.07389 (2019).

[6] Fake News Detection Using Machine Learning Approaches To cite this article: Z Khanam et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1099 012040.

[7] Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R. (2017). Automatic detection of fake news. arXiv preprint arXiv:1708.07104.

[8] Looijenga, M. S. "The Detection of Fake Messages using Machine Learning." 29 Twente Student Conference on IT, Jun. 6th, 2018, Enschede, The Netherlands. Netherlands: y.utwente.nl. 2018.

[9] Ahmed, Alim Al Ayub, et al. "Detecting fake news using machine learning: A systematic literature review." arXiv preprint arXiv:2102.04458 (2021).

[10] Mahmud, Fahim Belal, et al. "A comparative analysis of Graph Neural Networks and commonly used machine learning algorithms on fake news detection." 2022 7th International Conference on Data Science and Machine Learning Applications (CDMA). IEEE, 2022.

[11] E. C. Tandoc Jr et al. "Defining fake news a typology of scholarly definitions". Digital Journalism, 1–17. 2017.

[12] Gilda, S. "Evaluating machine learning algorithms for fake news detection." 15th Student Conference on Research and Development (SCOReD) (pp. 110-115). IEEE. 2017.

[13] Supanya Aphiwongsophon et al. " Detecting Fake News with Machine Learning Method." 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). Chiang Rai, Thailand, Thailand: IEEE . 2018.

[14] Abdullah-All-Tanvir, Mahir, E. M., Akhter S., Huq, M. R. (2019). Detecting Fake News using Machine Learning and Deep Learning Algorithms. 7th International Conference on Smart Computing Communications (ICSCC), Sarawak, Malaysia, Malaysia, 2019, pp.1-5, https://doi.org/10.1109/ICSCC.2019.8843612

[15] Granik, M., Mesyura, V. (2017). Fake news detection using naive Bayes classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, pp.900-903, https://doi.org/10.1109/UKRCON.2017.8100379

[16] S. Helmstetter, and H. Paulheim. "Weakly supervised learning for fake news detection on Twitter." 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2018.

[17] Roy, Arjun, et al. "A deep ensemble framework for fake news detection and classification." arXiv preprint arXiv:1811.04670 (2018).

[18] T. H. Borkar and T. Ahuja, "Comparative Study of Supervised Learning Algorithms for Fake News Classification," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:33 UTC from IEEE Xplore. Restrictions apply.

Comparative Analysis of ResNet and DenseNet for Differential Cryptanalysis of SPECK 32/64 Lightweight Block Cipher

Ayan Sajwan[1] and Girish Mishra[2]

[1] Delhi Technological University, Delhi - 110 042,India
ayansajwan2003@gmail.com,
[2] DRDO-Scientific Analysis Group, Delhi - 110 054, India
gmishratech28@gmail.com

Abstract. This research paper explores the vulnerabilities of the lightweight block cipher SPECK 32/64 through the application of differential analysis and deep learning techniques. The primary objectives of the study are to investigate the cipher's weaknesses and to compare the effectiveness of ResNet as used by Aron Gohr at Crypto2019 and DenseNet . The methodology involves conducting an analysis of differential characteristics to identify potential weaknesses in the cipher's structure. Experimental results and analysis demonstrate the efficacy of both approaches in compromising the security of SPECK 32/64.

Keywords: Differential Cryptanalysis, Deep Learning, Speck, ResNet, DenseNet

1 Introduction

Cryptography is the technique of converting data into an incomprehensible form known as cipher text. It is done by using mathematical principles and algorithms. This crucial sector ensures the security and privacy of modern digital communications and data storage. Throughout history, critical information ranging from military communications[1] to commercial transactions[2] have been protected via cryptographic processes.

Over the centuries, cryptography has been an art practised by many who have invented techniques to meet some of the information security requirements. The previous two decades have seen the field evolve from an art to a science[3].

Data secrecy,integrity and authenticity are the main goals of cryptography. Confidentiality ensures that only authorised personnel may access the information. Integrity ensures that the data is unchanged throughout transmission or storage. Authentication makes sure that only reliable sources are sharing information

With the aid of encryption algorithms[4], cryptography secures data and communication. This is achieved by converting plaintext into ciphertext, a form

that cannot be deciphered. On the other side, Cryptanalysis is the science of analysing cryptographic systems to find vulnerabilities. These weaknesses can be exploited to obtain the original plaintext or encryption keys. Lightweight block ciphers serve a critical role in cryptography in situations where computational resources are constrained. These ciphers' low computational and memory overhead makes them ideal for secure, effective encryption[5].

The study of decrypting cryptographic methods, or cryptanalysis, is a crucial area for maintaining the security of encryption systems. It involves investigating the mathematical features and design choices to find the flaws in the cipher . Understanding these flaws allows cryptanalysts to create more successful attacks. This helps in increasing the security of the cryptographic systems.

Numerous fields, including cryptanalysis[6], have seen the emergence of machine learning and deep learning as highly effective tools. These methods make use of the models' computational capabilities. They can automatically identify patterns, detect features, and generate predictions.Machine learning techniques can be used in the context of cryptanalysis as well[7]. It can be used to analyse and categorise cryptographic data, such as ciphertexts, plaintexts, or encryption keys.

Deep learning, a branch of machine learning, has achieved outstanding results in a number of fields. Ranging from speech recognition[8], computer vision[9] to natural language processing[10]. Deep neural networks are able to learn complex data representations and identify deep correlations.
In this paper, we investigate the use of deep learning methods, more specifically DenseNet and ResNet. They are used for differential cryptanalysis on the lightweight block cipher SPECK 32/64. We intend to compare the effectiveness of ResNet and DenseNet. It ultimately helps us in understanding of the security of the SPECK 32/64 cipher by utilising the expressive potential of deep neural networks.

2 SPECK 32/64 Overview

SPECK is a lightweight block cipher developed by the National Security Agency (NSA). It was a part of the Lightweight Cryptography Project of the NSA. Here, the term "lightweight" refers to cryptographic algorithms that are designed for efficient operation and low resource consumption.

This qualifies them for usage in environments with limited resources such as Internet of Things devices[11], wireless sensor networks[12], and embedded systems[13].

The SPECK family consists of a variety of block and key sizes. The block is made up of 2 words, and is of the form $2n$. Here, n is the size of the word which may be 16, 24, 32, 48 or 64 bits. The key size(k) is mn bits. The key contains 2,3 or 4 words depending on the variant. Hence, the SPECK family is of the form

SPECK 2n/mn and has ten variants. We use SPECK 32/64 in our work, which denotes 2 words of 16-bits each and 4 keys of 16-bits each as well.

The SPECK family cipher is made up of a Feistel network structure[14]. In this network, the input block is split into two equally sized halves. Encryption rounds are then performed on these halves using a subkey. A different subkey is derived for each round. The number of rounds differ in each variant of the family.

The round function of SPECK 32/64 uses a range of bitwise operations for its cryptographic operations. It includes rotation, XOR, and modular addition, to induce confusion and diffusion features, assuring the security of the cipher.

2.1 Round Function

Speck's round function is very simple.It is an ARX structure, which means it is made out of the fundamental functions of modular addition (mod 2^k), bitwise rotation, and bitwise addition. They are denoted by $\boxplus$, $\gg$ and $\oplus$ respectively. SPECK n/m represents Speck with n bit block size and m bit key size. It produces the next round state $(\mathbf{L_{i+1}}, \mathbf{R_{i+1}})$ with an input k-bit subkey $\mathbf{K}$ and the current cipher state consisting of two k-bit words $(\mathbf{L_i}, \mathbf{R_i})$. The algorithm is as follows:

$$\mathbf{L_{i+1}} = ((\mathbf{L_i} \gg \alpha) \boxplus \mathbf{R_i}) \oplus \mathbf{K}$$

$$\mathbf{R_{i+1}} = (\mathbf{R_i} \ll \beta) \oplus \mathbf{L_{i_i+1}}$$

The values of α and β are constant: $(\alpha = 7, \beta = 2)$ for Speck32/64 and $(\alpha = 8, \beta = 3)$ for other members of the Speck family. The cipher text output is generated by applying the round function on the plain text input for 22 rounds in the case of Speck 32/64. However, we refer to round reduced speck in this paper. The key used in each round is generated from a master key by applying a key schedule. The key schedule depends on the member of the Speck family, we refer to Beaulieu et al[15] in this paper for the key scheduling.

3 ResNet and DenseNet Architectures

3.1 ResNet

ResNet or Residual Network is a powerful deep learning architecture first published by Kaiming[16]. Resnet is intended to address the issue of disappearing gradients[17] in very deep neural networks. It does so, by incorporating residual connections[18] or skip connections, which enable the building of deeper and more precise models.

ResNet is composed of many residual blocks or towers that are layered on top of each other and contain a sequence of convolutions and a skip connection. The skip connection is added to a block's output and then passed on to the following block. This helps in reducing the vanishing gradient problem and allows for better model training. Figure 2 depicts the working of a skip connection.

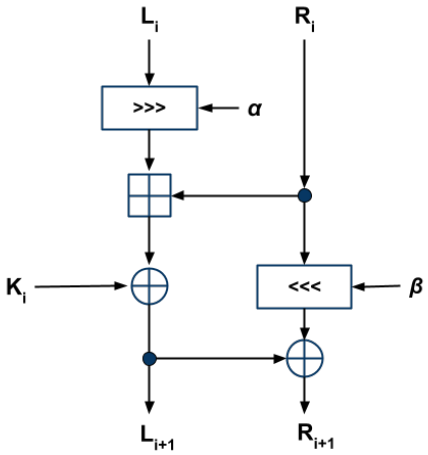


Fig. 1: General round of Speck

ResNet Structure The ResNet structure used in this work is the one used by Gohr in 2019[19]. It consists of a residual tower of depth ten, having a two layer convolutional network. The convolutional network has 32 filters.

First, Convolution is applied followed by a Batch normalization[20] for faster and stable training . It is followed by a Rectified Linear Unit layer[21] which introduces Non-linearity to the model. Then a skip/jump connection at the end adds the output of the final rectifier layer of the block to the convolutional block's input and forwards the result to the next block.

The initial layer is a bit-sliced 1 Dimensional Convolution with 32 output channels, which is followed by Batch normalization. Finally, a Rectified Linear Unit is applied to the preceding layer's output. The final result is a $32x16$ matrix that is fed into the depth-10 Residual Tower.

Finally, the data is flattened and transmitted to the prediction layer. This final layer consists of two densely linked hidden layers of 64 units each, followed by batch normalization, a Rectified Linear unit, and sigmoid activation for a single output head.

3.2 DenseNet

DenseNet is made up of Dense blocks and transition layers. DenseNet, which stands for "Densely Connected Networks" is a deep learning architecture designed by Gao Huang et al. originally published in their paper[22] in 2017.

DenseNet, like ResNet, aims to solve the vanishing gradient problem by maximising feature reuse. DenseNet introduces dense connections between layers and

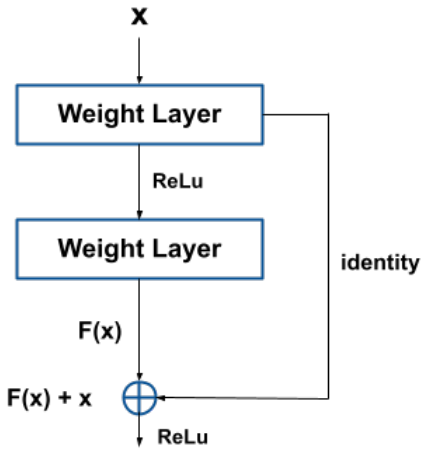


Fig. 2: Skip Connection

blocks as opposed to traditional neural networks, which connect layers sequentially. Unlike ResNet, which utilises an additive approach of adding previous layer output to subsequent layers, DenseNet uses all past outputs as input for future layers. As a result, each layer is directly linked to all the following layers.

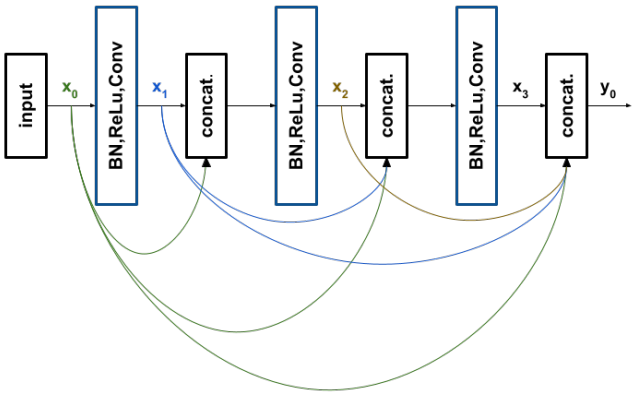


Fig. 3: Dense Connection

DenseNet Structure The first layer, like the one in the Resnet model, is a bit-sliced 1 Dimensional Convolution with 32 output channels.It is followed by

Batch normalisation, and lastly a Rectified Linear Unit is applied to the output of the preceding layer. The resulting 32x16 matrix is passed into the Dense Network.

DenseNet is made up of Dense blocks and transition blocks.The dense block has a depth-8 and is made up of two layers of 1D-convolution, Batch normalisation, and a Rectified linear unit layer. The convolution layer is made up of 64 filters and a kernel with a size of 3.Finally, the output is concatenated with the layer's input and handed on to the next layer.This occurs eight times since the depth is eight.

To restrict the amount of feature maps and minimise spatial dimensions, transition layers are inserted between dense blocks. The transition layer is made up of 1D convolution with 32 filters, batch normalisation, a ReLu layer, and 1D average pooling. The transition layer's output is subsequently passed on to the following dense block.

The dense block and transition layers are now merged with a depth of 2. This indicates alternating dense block and transition layer, followed by a final dense block. The final result is an overall structure of three dense blocks and two transition layers.

Finally, the data is flattened and transmitted to the prediction layer. This prediction layer consists of two dense hidden layers of 64 units each, followed by batch normalisation, Rectified linear unit, and sigmoid activation, similar to ResNet.

3.3 Input Data

Input data: Input consists of a pair of cipher texts (C_0, C_1). They are transformed into a 4x16 matrix with each row consisting of a word of the ciphertext. This way the data consists of four 16-bit words and therefore the input layer has 64 units. This input data is then passed into the ResNet and DenseNet architecture.

4 Experimental Setup and Methodology

4.1 Data Generation

The data generation methodology used is similar to the one used by Aron Gohr in 2019. A random number generator is used to create evenly distributed keys K_i and plain text pairings P_i with the input difference $\Delta = 0x0040/0000$, along with a vector of binary-valued real/random labels Y_i. If Y_i is set(=1), the plain text pair P_i is encrypted for $\mathbf{k}$ rounds to create training or validation data for k-round Speck, and if not, the second plain text in the pair is changed to a newly created random plain text. This way we have cipher texts belonging to 2 classes: Chosen Input difference ($\mathbf{Y} = 1$) and random input difference ($\mathbf{Y} = 0$). As a result we have 10^6 samples for our dataset for training and validation.

4.2 Training and Testing Procedure:

The data set of 10^6 samples is used for training in batches of 5000 and run for 20 epochs as opposed to 200 epochs by Gohr .
Mean Square Error (MSE)[23] loss is used with L2 weights regularization using the Adam algorithm[24] for optimization. This is a down scaled version of Gohr's experiment in which he used 10^7 samples for training for 200 epochs.

Testing data also contains a set of 10^6 samples with 2 classes. One of the chosen input difference and the other of random input difference. Table 1. provides the list of hyper-paramaters used in training with their values.

Hyper-parameters	**values**
Sample Size	10^6
Batch Size	5000
Epochs	20
Encryption Rounds	5,6,7,8
Optimizer	Adam
Loss function	MSE loss
Cyclic Learning Rate	0.002-0.0001

Table 1: Hyper-parameters for training of model

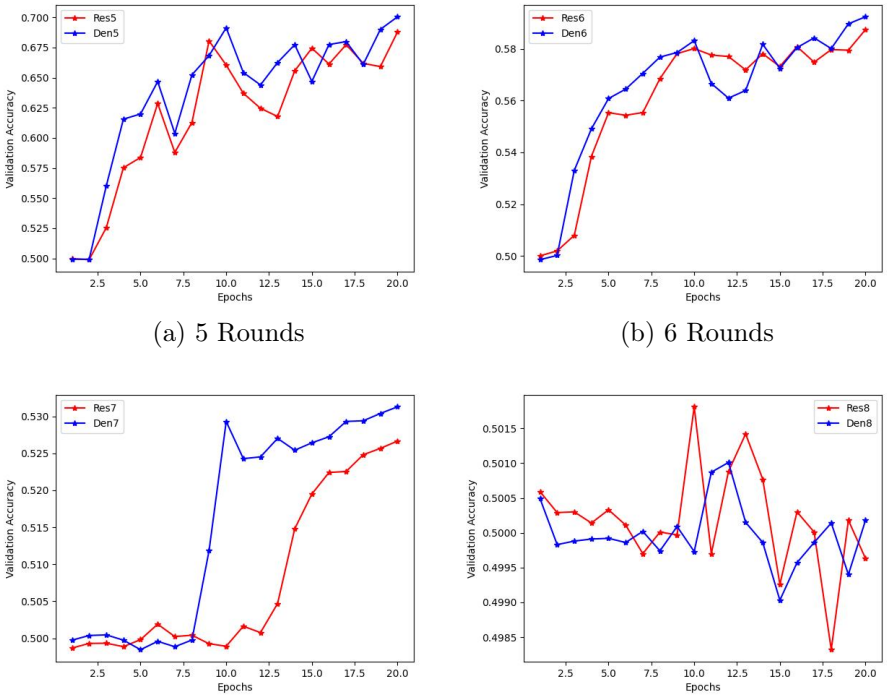
5 Results and Analysis:

In this section, we present the findings of our experiments on the differential cryptanalysis of the round reduced (rounds 5,6,7,8) SPECK 32/64 lightweight block cipher using the ResNet and DenseNet architectures. R5, R6, R7 & R8 refers to the ResNet architecture for rounds 5, 6, 7 and 8 respectively, and similarly D5, D6, D7 & D8 refers to the DenseNet architecture. Table 2 depicts the training and validation accuracy for both the models.

Rounds	**R (ResNet)**		**D (DenseNet)**	
	Training	Validation	Training	Validation
5	0.9332	0.6779	0.9309	**0.7005**
6	0.7952	0.5874	0.7917	**0.5923**
7	0.6096	0.5267	0.6053	**0.5313**
8	0.5012	0.4996	0.4998	0.5002

Table 2: Training and Validation accuracy for ResNet and DenseNet models

The DenseNet model achieved slightly better validation accuracy than the ResNet model for rounds 5, 6 and 7. For round 8, both the models failed to give a prominent result since the models could not learn an accurate pattern. Figure 4 below shows the comparison of accuracy for both the models.



(a) 5 Rounds (b) 6 Rounds

(c) 7 Rounds (d) 8 Rounds

Fig. 4: Validation Accuracy comparison

6 Conclusions

In this paper, we compared the ResNet and DenseNet architectures for differential cryptanalysis of the SPECK 32/64 lightweight block cipher. Our analysis attempted to evaluate their accuracy in deciphering the cipher's complicated differential behaviour.

According to our findings, the DenseNet architecture outperforms the ResNet architecture marginally. DenseNet achieved slightly higher predictions of ciphertext differences in the context of differential cryptanalysis. However, neither model produced a satisfactory result for an 8-round (or higher rounds) encryption cipher.

As of now, this work does not include a key retrieval approach. At last, our findings highlight the importance of architecture selection in differential crypt-

analysis. The observed accuracy improvements of DenseNet support its use in scenarios requiring lightweight block cipher analysis.

Acknowledgements

The Authors would like to show their sincere gratitude to the Scientific Analysis group (SAG),Defense Research and Development Organization (DRDO) for their invaluable support and collaboration throughout the course of this research. The authors would also like to thank Delhi Technological University (DTU),India for providing the opportunity to work in the field

References

1. Doukas, Nikolaos, and Nikolaos V. Karadimas. "A blind source separation based cryptography scheme for mobile military communication applications." WSEAS Trans. Commun 7.12 (2008): 1235-1245.
2. Lamprecht, C., et al. "Investigating the efficiency of cryptographic algorithms in online transactions." International Journal of Simulation: Systems, Science & Technology 7.2 (2006): 63-75.
3. Menezes, Alfred J., Paul C. Van Oorschot, and Scott A. Vanstone. Handbook of applied cryptography. CRC press, 2018.
4. Mahajan, Prerna, and Abhishek Sachdeva. "A study of encryption algorithms AES, DES and RSA for security." Global Journal of Computer Science and Technology 13.15 (2013): 15-22.
5. Hatzivasilis, G., Fysarakis, K., Papaefstathiou, I. et al. A review of lightweight block ciphers. J Cryptogr Eng 8, 141–184 (2018). https://doi.org/10.1007/s13389-017-0160-y
6. C. de Canniere, A. Biryukov and B. Preneel, "An introduction to Block Cipher Cryptanalysis," in Proceedings of the IEEE, vol. 94, no. 2, pp. 346-356, Feb. 2006, doi: 10.1109/JPROC.2005.862300.
7. Benamira, A., Gerault, D., Peyrin, T., Tan, Q.Q. (2021). A Deeper Look at Machine Learning-Based Cryptanalysis. In: Canteaut, A., Standaert, FX. (eds) Advances in Cryptology – EUROCRYPT 2021. EUROCRYPT 2021. Lecture Notes in Computer Science(), vol 12696. Springer, Cham. https://doi.org/10.1007/978-3-030-77870-5_28
8. L. Deng, G. Hinton and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: an overview," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 2013, pp. 8599-8603, doi:10.1109/ICASSP.2013.6639344.
9. Q. Wu, Y. Liu, Q. Li, S. Jin and F. Li, "The application of deep learning in computer vision," 2017 Chinese Automation Congress (CAC), Jinan, China, 2017, pp. 6522-6527, doi: 10.1109/CAC.2017.8243952
10. D. W. Otter, J. R. Medina and J. K. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 2, pp. 604-624, Feb. 2021, doi: 10.1109/TNNLS.2020.2979670.
11. Dinu, D., Corre, Y.L., Khovratovich, D. et al. Triathlon of lightweight block ciphers for the Internet of things. J Cryptogr Eng 9, 283–302 (2019). https://doi.org/10.1007/s13389-018-0193-x

12. M. Cazorla, K. Marquet and M. Minier, "Survey and benchmark of lightweight block ciphers for wireless sensor networks," 2013 International Conference on Security and Cryptography (SECRYPT), Reykjavik, Iceland, 2013, pp. 1-6.
13. Manifavas, C., Hatzivasilis, G., Fysarakis, K., Rantos, K. (2014). Lightweight Cryptography for Embedded Systems – A Comparative Analysis. In: Garcia-Alfaro, J., Lioudakis, G., Cuppens-Boulahia, N., Foley, S., Fitzgerald, W. (eds) Data Privacy Management and Autonomous Spontaneous Security. DPM SETOP 2013 2013. Lecture Notes in Computer Science(), vol 8247. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-54568-9_21
14. Nyberg, K. (1996). Generalized Feistel networks. In: Kim, K., Matsumoto, T. (eds) Advances in Cryptology — ASIACRYPT '96. ASIACRYPT 1996. Lecture Notes in Computer Science, vol 1163. Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0034838
15. Beaulieu, Ray, et al. "The SIMON and SPECK families of lightweight block ciphers." cryptology eprint archive (2013).
16. He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
17. H. Li, J. Li, X. Guan, B. Liang, Y. Lai and X. Luo, "Research on Overfitting of Deep Learning," 2019 15th International Conference on Computational Intelligence and Security (CIS), Macao, China, 2019, pp. 78-81, doi: 10.1109/CIS.2019.00025.
18. Jastrzębski, S., Arpit, D., Ballas, N., Verma, V., Che, T., & Bengio, Y. (2017). Residual connections encourage iterative inference. arXiv preprint arXiv:1710.04773.
19. Gohr, Aron. "Improving attacks on round-reduced speck32/64 using deep learning." Advances in Cryptology–CRYPTO 2019: 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18–22, 2019, Proceedings, Part II 39. Springer International Publishing, 2019.
20. Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International conference on machine learning. pmlr, 2015.
21. Agarap, Abien Fred. "Deep learning using rectified linear units (relu)." arXiv preprint arXiv:1803.08375 (2018).
22. Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
23. Toro-Vizcarrondo, Carlos, and T. Dudley Wallace. "A test of the mean square error criterion for restrictions in linear regression." Journal of the American Statistical Association 63.322 (1968): 558-572.
24. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

Proceedings of the Second International Conference on Edge Computing and Applications (ICECAA 2023)
IEEE Xplore Part Number: CFP23BV8-ART; ISBN: 979-8-3503-4757-9

Comparison of Fuzzy-C-Means Method with K-Means Algorithm and Self Organizing Maps using Image Quantization

Rishabh Chauhan
Department of Applied Mathematics
Delhi Technological University
Delhi
rishabhchauhan_2k19mc104@dtu.ac.in

Rajan Yadav
Department of Applied Mathematics
Delhi Technological University
Delhi
rajanyadav_2k19mc100@dtu.ac.in

Rajesh Kumar Gupta
Department of Applied Mathematics
Delhi Technological University
Delhi
rajeshkumargupta_2k19mc102@dtu.ac.in

2023 2nd International Conference on Edge Computing and Applications (ICECAA) | 979-8-3503-4757-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICECAA58104.2023.10212128

Abstract— There are several applications for the crucial colour quantization process in the disciplines of graphics and image processing. The majority of quantization methods are built on data clustering algorithms. Hard c-means (k-means) clustering methods have recently been found to be effective in this field. Other investigations came to similar conclusions with the fuzzy c-means approach and Self-Organizing Maps (SOM). The three different clustering techniques weren't explicitly compared in any of these experiments, which is interesting. In this study, we develop quick and accurate variations of the hard and fuzzy c-means algorithms coupled with SOM and a variety of starting strategies, and then compare the resulting quantizers on a wide range of images. The findings show that images produced by K-means and SOM had a little lower PSNR value than those produced by the fuzzy c-means algorithm. Due to their widespread use in this sector, along with Hierarchical clustering and Density Based-Spatial Clustering of Applications with Noise (DBSCAN), we have picked FCM, K-means, and SOM for comparison. Although we haven't discussed about hierarchical clustering and DBSCAN because they are space and time costly algorithms.

Keywords—Fuzzy C Means (FCM), Self-Organizing Maps(SOM)

I. INTRODUCTION

Over the past two decades, true-color photographs have proliferated. There may be hundreds of thousands of colours in a typical true-color image, making it difficult to display, store, transmit, process, and analyse such an image. A finite number of discrete levels are used to approximate the continuous colour space of an image during the quantization process. The quantization inaccuracies introduced by this approximation, however, can cause apparent artifacts and a decline in image quality.

Image quantization, therefore, aims to reduce quantization mistakes while maintaining the visual quality of the image. Because it lowers the image's storage and transmission needs, image quantization is a crucial step in the processing of images. The division of an image into a range of distinct colours or shades of grey is known as quantization. Clustering techniques are frequently used for this, which classify the image's pixels into clusters based on how similar they are.

Three widely utilized clustering methods that have been extensively employed for picture quantization include fuzzy C-Means, k-means, and self-organizing maps. Over the years, a number of picture quantization techniques—from straightforward uniform quantization to cutting-edge ones based on machine learning—have been developed. Depending on the exact application and the required trade-off between image quality and storage/bandwidth needs, each approach has pros and cons. To achieve the ideal balance between image quality and resource efficiency, it is essential to choose a suitable image quantization method.

According to experimental findings, fuzzy c-means only marginally outperforms other purpose-built colour quantization systems while rapid fuzzy clustering algorithms yield quantization results that are on par with the full classical fuzzy c-means technique. In this study, we review various colour quantization methods. We demonstrate that colour quantization, where colour cluster boundaries are not well specified, is a situation in which the fuzzy membership paradigm is particularly well suited.

We provide a thorough analysis of picture quantization methods and assess the compression efficiency of the quantized images. To do this, we drew on a number of libraries, including numpy, matplotlib, pillow, OpenCV, sklearn, and minisom.

After picture quantization, we compared the compressed image sizes and determined the PSNR value to categorise the three algorithms.

II. K MEANS CLUSTERING

An unsupervised algorithm called K-Means places each pixel in a single cluster based on how near it is to the centroid. K-Means is computationally efficient and provides good results for images with high contrast. When we have unlabeled data—data without clearly defined categories or groups—the algorithm is applied. The K-means algorithm consists of two sections. After identifying the k centroid, each point is subsequently relocated to the cluster that has the closest centroid to the relevant data point.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:21:29 UTC from IEEE Xplore. Restrictions apply.

The Euclidean distance is used to determine which centroid is nearest. After grouping is complete, it updates each cluster's centroid and calculates a new Euclidean distance between each centre. The cluster points with the shortest Euclidean distance are assigned to each data point based on that centroid. The centroid and members (member objects) of each cluster in the partition serve as unique identifiers. The centroid of each cluster is the location from which all of the elements in the cluster are separated by the least amount of distance. The overall distance between every object and every cluster's centroid is decreased. K-Means may, however, yield subpar results for photos with low contrast or intricate texture patterns. Grayscale and coloured images can be quantized using K-Means.

Consider an image with an x-y resolution that needs to be grouped into several groups. The following is the k-means clustering algorithm.:

1. Set the cluster's centre randomly and initialize k value.
2. Find the Euclidean distance (d) between the centre of the image and each individual pixel using the formulae shown below.

$$d = || p(x,y) - C_k || \quad (1)$$

$$d = \sqrt{(x - C_x)^2 + (y - C_y)^2} \quad (2)$$

where,
p(x,y) = point with co-ordinate (x,y).
x = x component of pixel.
y = y component of pixel.
Cx = x component of cluster centroid.
Cy = y component of cluster centroid.

3. Determine the closest location for each pixel's centre depending on the distance d.
4. Once all of the pixels have been assigned, recalculate the new location of the centre using the relation given below.

$$C_k = \frac{1}{k} \sum_{y \in C_k} \sum_{x \in C_k} p(x,y) \quad (3)$$

where,
p(x,y) = pixel with components (x,y).
x = x component of pixel.
y = y component of pixel.
Ck = cluster centroid.
K = number of clusters

5. Till the tolerance or error value is reached, keep going with the process.
6. Utilise the grouped pixels to generate new images.

III. Fuzzy C Means Clustering

An algorithm for clustering called fuzzy c-means (FCM) divides a dataset into a predetermined number of clusters. The soft clustering technique Fuzzy C-Means (FCM) divides each pixel into numerous clusters according to the degree of membership. FCM has been widely used in image processing because it is computationally efficient and provides better results than traditional hard clustering algorithms like K-Means. FCM can be used to quantize both grayscale and color images. However, FCM has some limitations, such as sensitivity to initialization, and it can produce noisy results in images with low contrast.

The algorithm involves the following step:
1. Initialization: Determine the quantity of clusters, then initialise the cluster centroids at random.

2. Assigning Membership Values: Assign each data point a membership value indicating the degree of belongingness to each cluster. The membership values are computed using the following equation:

$$U_{ij} = \frac{1}{\sum_{k=1}^{C} (\frac{d_{ij}}{d_{ik}})^{2/(m-1)}} \quad (4)$$

where
U_{ij} = *membership value of data point 'i' to cluster 'j'*
d_{ij} = *distance between data point 'i' and cluster centroid 'j'*
d_{ik} = *distance between data point 'i' and cluster centroid 'k'*
c = *total number of clusters*
m = *parameter that regulates how fuzzy the clustering is.*

3. Updating Cluster Centroids: Update the cluster centroids using the membership values. The updated centroids are created using the weighted average of the data points, where the weight is the membership value of each data point.

4. Repeat Steps 2-3 until convergence: Iterate the assigning membership values and updating cluster centroids until the membership values converge to a stable solution.

IV. Self Organizing Maps

SOM is a clustering algorithm for neural networks that converts high-dimensional input data to a grid of low-dimensional neurons. SOM can disclose the image's underlying structure while preserving the topology of the input space. As a result, it is frequently utilised in image processing. SOM is particularly useful for color image quantization because it can capture the color distribution of the image. However, SOM is computationally intensive and requires a large amount of memory for training.
The following is the step-by-step procedure of the SOM algorithm:

1. The SOM grid, which comprises of a collection of neurons arranged in a 2D or 3D grid form, should be initialised.
2. Randomly initialize the weight vectors of each neuron in the grid to a small random value.
3. Select an input data point from the dataset.
4. Locate the neuron in the grid whose weight vector is closest to the input data point. The Best Matching Unit (BMU) is the name given to this neuron.
5. Use the following formula to adjust the BMU and its surrounding neurons' weights in the grid:

$$W_{ij} = W_{ij} + \alpha(t) * (X_{ik} - W_{ij}(old)) \quad (5)$$

where

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:21:29 UTC from IEEE Xplore. Restrictions apply.

W_{ij} = weight vector of neuron i,j
X_{ik} = input data point
$\alpha(t)$ = learning rate at time t
$h(C_i, C_{bmu})$ = neighborhood function that determines the influence of the BMU on the neighboring neurons,
C_i, and C_{bmu} = coordinates of neuron i,j and the BMU, respectively.

6. Repeat steps 3-5 for a fixed number of iterations or until convergence is reached.
7. Use the trained SOM grid to cluster the input data points based on the similarity of their weight vectors.

The SOM algorithm can reduce high-dimensional data to a lower-dimensional space while maintaining the input space's structure, making it particularly helpful for data visualization. Additionally, SOMs can be used for feature extraction and data compression tasks by using the weight vectors of the neurons as feature representations of the input data.

V. Implementation Methodology

Python has been used to build the self-organizing maps, fuzzy C-means, and K-means algorithms. The image datatype is handled by the packages Pillow, OpenCV, and matplotlib. For picture compression, we utilised "https://compressjpeg.com/".
1.Read image using an image library like OpenCV, Pillow.
2.Transform image into dataset containing data about pixels.
3.Train the algorithm using the dataset.
4. Assign colours to the pixels based on the centroid closest to them to form the quantized image.
5.Transform the dataset to a new image.

RESULT ANALYSIS FACTOR

PSNR: In order to assess how well a picture may be represented, the peak signal-to-noise ratio (PSNR) compares the maximum power of a picture to the maximum power of corrupting noise. To determine a picture's PSNR, it's critical to contrast it with the clearest, most ideal image possible.

$$PSNR = 10\log_{10}\left(\frac{(L-1)^2}{MSE}\right) = 20\log_{10}\left(\frac{(L-1)}{RMSE}\right) \quad (6)$$

Here,
L = number of maximum possible intensity levels in an image, with 0 being the smallest intensity level.
MSE (mean squared error), is described as follows:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(O(i,j) - D(i,j))^2 \quad (7)$$

where ,
O = original image's matrix data.
The matrix data of deteriorated photographs is represented by D.
m = count of rows of pixel
i = index of the row inside the image.
n = count of columns of pixel
j = index of that column inside the image.
RMSE = Root mean squared error.

Compression size and Percentage: Comparison of the size of the images before compression and after compression and the percentage reduction according to their relative original images is also a factor to be considered for classifying image quantization algorithms.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:21:29 UTC from IEEE Xplore. Restrictions apply.

FCM will give more better images than K-means because of its soft clustering nature which doesn't not allow colour changes to occur suddenly in the quantized image.

Comparison of images before and after compression

Table 1

Before Compression	Algorithms	After Compression
	None	
	K means	
	FCM	
	SOM	
	None	
	K means	
	FCM	
	SOM	

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:21:29 UTC from IEEE Xplore. Restrictions apply.

Comparison of performance of the algorithms

Table 2

Photo	Quantization algorithm used	Size before compression	% of size reduction after compression	Size after compression	PSNR value
	None	108 kB	-1%	106.92 kB	100 dB
	K means Clustering	184 kB	-45%	101.2 kB	29.3659 dB
	Fuzzy C Means	76 kB	-6%	71.44 kB	30.0327 dB
	Self-Organizing Maps	82.4kB	-7%	76.6 kB	28.38dB
	None	33 kB	-24%	25.08 kB	100 dB
	K means Clustering	49.6 kB	-47%	26.28 kB	30.13 dB
	Fuzzy C Means	21.9 kB	-10%	19.71 kB	31.6 dB
	Self-Organizing Maps	15.3 kB	-30%	10.71 kB	29.04 dB

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:21:29 UTC from IEEE Xplore. Restrictions apply.

VI. Results

The outcomes demonstrate that both algorithms generated high-quality images. However, in terms of execution time and computational complexity, the Fuzzy C-Means approach surpassed K-means. Fuzzy C-Means had a lower computational complexity than the K-means technique. Although the PSNR values for the two methods were similar, the Fuzzy C-Means approach outperformed the K-means technique by a little margin.

Time Complexity:

Time Complexity of all three algorithms depends upon different factors like image size, number of clusters, implementation of algorithm and hardware. But in general terms we can say that: -

K-means (least number of iterations) > FCM (number of iterations greater than K-means) > SOM (number of iterations greater than K-means and FCM both).

.

VII. Conclusion

In this paper, we compared three popular clustering algorithms, Fuzzy C-Means, K-Means, and Self-Organizing Maps, for image quantization. PSNR value of all three algorithms was almost the same, but fuzzy c means had slightly higher value than the rest. The effectiveness of these algorithms depends on the characteristics of the image being quantized. The compression size of the quantized images was reduced significantly in all three cases, while k means showing images rather greater in size than the other two algorithms.

Photo 1:

Original Compressed Image (106.92 kB) > K means quantized compressed image (101.2 kB) > Fuzzy C means quantized compressed image (71.44 kB) > SOM quantized compressed image (76.6 kB)

Photo 2:

Original Compressed Image (25.08 kB) > K means quantized compressed image (26.28 kB) > Fuzzy C means quantized compressed image (19.71 kB) > SOM quantized compressed image (10.71 kB). As a result, the clustering algorithm selection should be based on the particular needs of the image processing application.

References

[1] Advances In Research Of Fuzzy C-means Clustering Algorithm Chen Yanyun , Qiu Jianlin, Gu Xiang,Chen Jianping, Ji Dan, Chen Li
College of Computer Science and Technology, Nantong University, Nantong, China.

[2] The global k-means clustering algorithm Aristidis Likasa; ∗, Nikos Vlassisb, JakobJ. Verbeekb aDepartment of Computer Science, University of Ioannina, 45110 Ioannina, Greece bComputer Science Institute, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands.

[3] Self-organizing map Teuvo Kohonen, Samuel Kaski, Panu Somervuo, Krista Lagus, Merja Oja, Vesa Paatero.

[4] An Efficient Fuzzy C-Means Clustering Algorithm Ming-Chuan Hung and Don-Lin Yang Department of Information Engineering, Feng Chia University 100 Wenhwa Rd., Taichung, Taiwan.

[5] Image Compression using self-Organizing Maps C. Amerijckx, J.-D Legat & M Verleysen**.**

[6] A comparison of clustering algorithms applied to color image quantization P. Scheunders

[7] A Hardware-Efficient Vector Quantizer Based on Self-Organizing Map for High-Speed Image Compression Zunkai Huang 1,2,3 ID , Xiangyu Zhang 3 , Lei Chen 3 , Yongxin Zhu 1 , Fengwei An 3,*, Hui Wang 1,* and Songlin Feng

[8] Image quality metrics: PSNR Alain Horé MOIVRE, Département d'informatique, Faculté des sciences, Université de Sherbrooke Sherbrooke (Québec), Canada, J1K2R1 e-mail: alain.hore@usherbrooke.ca Djemel Ziou MOIVRE, Département d'informatique, Faculté des sciences, Université de Sherbrooke Sherbrooke (Québec), Canada, J1K2R1 e-mail: djemel.ziou@usherbrooke.ca

[9] Comparison and Optimization of Methods of Color Image Quantization Jean-Pierre BRAQUELAIRE and Luc BRUN LaBRI 351, cours de la liberation F-33405 Talence CEDEX

[10] Color Image Quantization: A Short Review and an Application with Artificial Bee Colony Algorithm Celal OZTURK∗ , Emrah HANCER, Dervis KARABOGA Department of Computer Engineering, Erciyes University, Kayseri, 38039, Turkey e-mail: celal@erciyes.edu.tr,emrahhancer@erciyes.edu.tr,karaboga@erciyes.edu.tr.

[11] Fuzzy c-means clustering with spatial information for image segmentation Keh-Shih Chuang a,*, Hong-Long Tzeng a,b, Sharon Chen a , Jay Wu a,b, Tzong-Jer Chen c a Department of Nuclear Science, National Tsing-Hua University, Hsinchu 30013 Taiwan b Health Physics Division, Institute of Nuclear Energy Research, Atomic Energy Council, Taiwan c Department of Medical Imaging Technology, Shu-Zen College of Medicine and Management, Kaohsiung, Taiwan.

[12] J. M. Barbalho, A. Duarte, D. Neto, J. A. F. Costa and M. L. A. Netto, "Hierarchical SOM applied to image compression," IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222), Washington, DC, USA, 2001, pp. 442-447 vol.1, doi: 10.1109/IJCNN.2001.939060.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:21:29 UTC from IEEE Xplore. Restrictions apply.

arXiv:2307.14691v1 [gr-qc] 27 Jul 2023

Constraints on the Parameters of Modified Chaplygin-Jacobi and Modified Chaplygin-Abel Gases in $f(T)$ Gravity Model

Himanshu Chaudhary,[1,2,3,*] Ujjal Debnath,[4,†] Tanusree Roy,[4,‡] Sayani Maity,[5,§] and G.Mustafa[6,7,¶]

[1] *Department of Applied Mathematics, Delhi Technological University, Delhi-110042, India,*
[2] *Pacif Institute of Cosmology and Selfology (PICS), Sagara, Sambalpur 768224, Odisha, India*
[3] *Department of Mathematics, Shyamlal College, University of Delhi, Delhi-110032, India,*
[4] *Department of Mathematics, Indian Institute of Engineering Science and Technology, Shibpur, Howrah-711 103, India,*
[5] *Department of Mathematics, Techno India Salt Lake, Sector-V, Kolkata-700 091, India.*
[6] *Department of Physics, Zhejiang Normal University, Jinhua 321004, Peoples Republic of China,*
[7] *New Uzbekistan University, Mustaqillik ave. 54, 100007 Tashkent, Uzbekistan,*

In this work, we explore the parameter constraints of two dark energy models, namely the Modified Chaplygin-Jacobi gas (MCJG) and Modified Chaplygin-Abel gas (MCAG), within the context of $f(T)$ gravity model in a non-flat Friedmann-Lemaître-Robertson-Walker (FLRW) Universe. The Universe is assumed to be composed of three main components: radiation, dark matter, and dark energy. Our investigation involves comparing the equation of state for the MCJG and MCAG dark energy models with the equation of state derived from the $f(T)$ gravity model. To derive constraints for the dark energy and $f(T)$ gravity models, we use recent astronomical datasets, including $H(z)$ data, type Ia supernovae observations, Gamma Ray Bursts data, quasar data, and Baryon Acoustic Oscillation (BAO) measurements. These datasets serve as valuable sources of information to understand the behavior and properties of the Universe on cosmic scales. We present the reduced Hubble parameter in terms of observable parameters such as Ω_{r0} (density parameter of radiation), Ω_{m0} (density parameter of dark matter), Ω_{k0} (density parameter associated with spatial curvature), Ω_{CJ0} (density parameter of Modified Chaplygin-Jacobi gas), Ω_{CA0} (density parameter of Modified Chaplygin-Abel gas), and H_0 (present value of the Hubble parameter). We explore the cosmological evolution through various cosmic diagnostic parameters, including the deceleration parameter, Om diagnostic, and statefinder diagnostic pair $\{r, s\}$. These diagnostic parameters offer valuable insights into the expansion dynamics and the nature of dark energy in the Universe. We have also assessed the viability of the models using the information criteria. By comparing the results with the standard ΛCDM model, we have gained insights into the feasibility of our proposed models. Our aim is to shed light on the nature of dark energy and its connection to the $f(T)$ gravity model, and ultimately gain a deeper understanding of the underlying mechanisms driving the accelerated expansion of our Universe.

I. INTRODUCTION

Over the course of last decade numerous astronomical observations have been carried out to investigate the compatibility of theoretical cosmological models. Recent observations of the type Ia supernova (SNIa)[1–3], Baryon Acoustic Oscillation (BAO)[4, 5], cosmic microwave background (CMB)[6], WMAP [7, 8] have confirmed the most astounding result of modern cosmology that not only our Universe is expanding but it is passing through an era of accelerated expansion. Now, the standard Big Bang cosmology was unable to resolve this 'cosmic puzzle'; so the possible existence of some exotic matter, acclaimed as Dark Energy (DE), with a unique characteristic of negative pressure, has emerged which would obey $\rho + 3p < 0$ as the Friedmann equation reads

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) \tag{1}$$

and $\ddot{a} > 0$ for accelerated expansion. At first ΛCDM model was considered as the 'best fit'[9, 10], but fine-tuning problem and the coincidence problem could not be explained with this model [11–13]. To resolve these severe issues

* himanshuch1729@gmail.com
† ujjaldebnath@gmail.com
‡ tanusreeroy1995@gmail.com
§ sayani.maity88@gmail.com
¶ gmustafa3828@gmail.com

theorists suggested for modification of Einsteins field equations in two ways. One approach is the correction of the matter field that yields various dynamical dark energy models with varying equation of state such as quintessence, tachyon, k-essence, fermionic field, phantom, Chaplygin gas, DBI essenec[13, 14] etc. In search of a befitting candidate for DE, Chaplygin[15] originated a model called Chaplygin gas (CG), defined with a simple equation of state

$$p = -\frac{B}{\rho} \tag{2}$$

where B is considered as positive constant. This model unified dark matter and DE because it acts as dust-like matter at early stage and cosmological constant at later time. But one backdrop of this model is that it can not explain the era of the structure formation in the Universe [16, 17]. In this regard, the pure CG model is generalized to the generalized Chaplygin gas (GCG) model [18] with equation of state

$$p = -\frac{B}{\rho^{\alpha}}, \quad \text{with} \quad 0 < \alpha \leq 1. \tag{3}$$

In Ref.[19, 20] Benaoum proposed Modified Chaplygin Gas (MCG) model that obeys the EoS

$$p = A\rho - \frac{B}{\rho^{\alpha}} \tag{4}$$

where $A, B > 0$ and $0 \leq \alpha \leq 1$. For $A = 1/3$, the model exhibits a radiation era when assuming a minuscule scale factor. On the other hand, with an infinitely large scale factor, it acts as the ΛCDM model. In an intermediate stage, it behaves as pressure-less dust. Further along, the EoS was revised to derive other extensions like variable CG[21], variable modified CG[22], generalized cosmic Chaplygin gas (GCCG), the new variable modified Chaplygin gas (NVMCG)[23], viscous CG[24] etc. For better synchronization of the models to the observational data, constraints on different parameters of CG and MCG were investigated[25–30]. Using Jacobi's elliptic function, Villanueva formulated generalized Chaplygin-Jacobi gas correlated with the generalized Chaplygin scalar field[31–33] which can describe the inflationary stage and confronted the measurement recently released by the Planck 2015 data. Following that, Debnath derived the equation of states for Modified Chaplygin-Abel gas and Modified Chaplygin-Jacobi gas and explored the models' compatibility with observational data using AIC, BIC, and DIC model selection tools[34]. Accretion phenomena of Modified Chaplygin-Jacobi Gas and Modified Chaplygin-Abel Gas models into Schwarzschild Black Hole have been studied in ref [35] .

Another approach to explain the cosmic acceleration of the Universe is constructing gravitational modifications in the geometric part of the Einsteins field equations so that the additional degree(s) of freedom can act as a driving force in the accelerated expansion [36]. Mostly modified gravity theories are formed by extending the Einstein-Hilbert action i.e. by curvature formulation. Some of the widely used examples are $f(R)$ gravity [37, 38]; $f(G)$ gravity[39]; Horava-Lifshitz gravity[40]; Lovelock gravity[41] etc. Another suitable approach to gravitational modification is the extension of the action of equivalent torsional formulation of General Relativity (GR). The simplest modification is $f(T)$ gravity [42–44], developed from the Teleparallel Equivalent of General Relativity (TEGR) [45–50], where T is the Torsion scalar. The characteristic difference between TEGR with GR is that the Lagrangian of TEGR contains a curvature-less Weitzenbock connection whereas GR has a torsion-less Levi-Civita connection. The desirability of $f(T)$ gravity model over $f(R)$ gravity is the second-order field equation in contrast with fourth-order field equations in $f(R)$ gravity. In teleparallelism, torsion is formed from the product of the first derivatives of the tetrad having no second derivative terms. Thus the Lagrangian can simply be extended by just adding a function $f(T)$. Both the inflation phase [51–54] and acceleration phase[55–57] of the Universe have been studied under the framework of $f(T)$ gravity. As $f(T)$ gravity incorporates an unknown function, the viability of the model remains questionable from theoretical and observational perspectives depending on the function chosen. Using Monte-Carlo analysis, $f(T)$ gravity models have been investigated already[58]. Constraining the free parameters have been executed with data sources like Pantheon Supernovae sample[59], redshift-space distribution measurements [59] or solar system data[60]. Previous works in this domain have suggested that DE coupled with a modified gravity field can present fascinating results. For example, Jamil et al answered the cosmic coincidence problem by working on MCG in loop quantum gravity[61]. Park pointed out that additional terms like a^{-4}, a^{-2} contained in Friedmann equation in Horava-Lifshitz gravity can be responsible for cosmic acceleration[62]. Many other gravity theories like Chern-Simons gravity[63], Galileon gravity[64], Einstein-Aether gravity[65], Brans-Dicke theory[66], were combined with MCG to determine the observational constraints on the parameters. Motivated by this line of work, here we will review the observational

constraints of MCJG and MCAG parameters with association of various $f(T)$ gravity models. The free parameters of the $f(T)$ models have also been constrained with the help of different combinations of current observational data.

The flow of work of this paper is as follows: the basic equations of $f(T)$ gravity models is described in section II. In section III the viable $f(T)$ gravity model is presented. Section IV is devoted to investigate the constraints of the model parameters using various observational dataset comprising of $H(z)$, Pantheon, Gamma Ray Bursts, quasars, and BAO measurements. Comparison of observational and theoretical predictions of model parameters are done via $H(z)$ measurements and distance modulus $\mu(z)$ in section V. For cosmographic analysis the deceleration parameter, jerk parameter are derived in section VI. The statefinder diagnostic OM diagnostic are analyzed graphically in section VII and VIII respectively. In section IX, we present the information criteria for both the models. Finally in section X, the results are summarized and the main conclusion is exposed.

II. BASIC EQUATIONS IN $f(T)$ GRAVITY

One of the most intriguing generalization of Einstein's general relativity in the context of explanation of the observed cosmic acceleration of the Universe is the $f(T)$ theories of gravity. Originally teleparallelism was introduced by Einstein in the process of unification of gravity and electromagnetism. Unlike the General theory of Relativity that uses the torsionless Levi-Civita connection, this theory exploits the curvatureless Weitzenb¨ock connection that has only torsion. The geometry of a non-flat Friedmann-Lemaître-Robertson-Walker (FLRW) model of the Universe is described by the metric

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + sin^2\theta d\phi^2\right)\right] \tag{5}$$

where $a(t)$ is the scale factor and $k(= 0, -1, +1)$ is the curvature scalar which represents flat $(k = 0)$, open $(k = 1)$, and closed $(k = -1)$ model of the Universe. The relation of orthonormal tetrad components $\mathbf{e_A}(\mathbf{x}^\mu)$ relate and the metric is

$$g_{\mu\nu} = \eta_{AB} e^A_\mu(x) e^B_\nu(x), \tag{6}$$

where $\mathbf{e_A}(\mathbf{x}^\mu) = \mathbf{e}^\mathbf{A}_\mu \partial_\mu$, the indices A, B run over 0, 1, 2, 3 for the tangent space of the manifold, the coordinate indices on the manifold μ, ν take values 0, 1, 2, 3 and i, j are the spatial indices. In this formalism the torsion tensor takes the form

$$T^\lambda_{\mu\nu} = \Gamma^{\mathbf{W}^\lambda}_{\nu\mu} - \Gamma^{\mathbf{W}^\lambda}_{\mu\nu} = e^\lambda_A(\partial_\mu e^A_\nu - \partial_\nu e^A_\mu). \tag{7}$$

The contorsion tensor reads

$$K^{\mu\nu}_\rho = -\frac{1}{2}(T^{\mu\nu}_\rho - T^{\nu\mu}_\rho - T^{\mu\nu}_\rho) \tag{8}$$

From torsion tensor and contorsion tensor a new quantity is defined

$$S^{\mu\nu}_\rho = \frac{1}{2}(K^{\mu\nu}_\rho + \delta^\mu_\rho T^{\alpha\nu}_\alpha - \delta^\nu_\rho T^{\alpha\mu}_\alpha) \tag{9}$$

The Lagrangian of the Telepaarallel gravity is the torsion scalar T defined as follows

$$T = S^{\mu\nu}_\rho T^\rho_{\mu\nu}. \tag{10}$$

Following the footsteps of $f(R)$ gravity, $f(T)$ gravity is formed by extending the torsion scalar T to $T + f(T)$ where $f(T)$ is a differentiable function of T. The action for $f(T)$ gravity can be written as[43, 44, 67, 68]

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{-g}\,[T + f(T) + \mathcal{L}_m] \tag{11}$$

where $\mathcal{L}_m$ is denoted as the Lagrangian for matter content. TEGR model is retrieved as $f(T) \to 0$. It exhibits the ΛCDM model as $f(T)$ approaches a constant value Λ.

In $f(T)$ gravity, the Friedmann equations are written for non-flat FLRW Universe [69, 70]

$$\left(H^2+\frac{k}{a^2}\right)+2H^2f'(T)+\frac{f(T)}{6}=\frac{8\pi G}{3}\,\rho \tag{12}$$

$$\left(\dot{H}-\frac{k}{a^2}\right)\left(1+f'(T)+12H^2f''(T)\right)=-4\pi G(\rho+p) \tag{13}$$

where, $H=\dot{a}/a$ is the Hubble parameter and the torsion scalar $T=-6\left(H^2+\frac{k}{a^2}\right)$ with prime and dot denote the derivatives w.r.t. T and t respectively. We assume that the total energy density and pressure of the fluid that fills the Universe are ρ and p, respectively. So the energy conservation equation takes the form

$$\dot{\rho}+3H(\rho+p)=0 \tag{14}$$

We start with the prediction that the Universe is composed of matter content comprising radiation, dark matter (DM), and dark energy (DE). So, ρ and p consist of densities and pressures of radiation, DM and DE. So $\rho=\rho_r+\rho_m+\rho_d$ and $p=p_r+p_m+p_d$. Now assume that radiation, DM, and DE follow the conservation equation separately so that

$$\dot{\rho}_r+3H(\rho_r+p_r)=0, \tag{15}$$

$$\dot{\rho}_m+3H(\rho_m+p_m)=0 \tag{16}$$

and

$$\dot{\rho}_d+3H(\rho_d+p_d)=0 \tag{17}$$

The relation between energy density and pressure for radiation is, $p_r=\frac{1}{3}\rho_r$. Solving equation (15) we obtain $\rho_r=\rho_{r0}a^{-4}$. Since pressure for the DM can be considered negligible (i.e., $p_m=0$), equation (16) gives $\rho_m=\rho_{m0}a^{-3}$.
• **MCJG :** Now consider the DE in the form of modified Chaplygin-Jacobi gas (MCJG) whose equation of state is given by [71]

$$p_d=[(2K-1)(1+A)-1]\rho_d-\frac{KB}{\rho_d^{\alpha}}+\frac{(1-K)(1+A)^2}{B}\,\rho_d^{2+\alpha} \tag{18}$$

where $K\in[0,1]$, $\alpha\in[0,1]$, $A(>0)$ and $B(>0)$ are constants. Putting the above expression in (17), we obtain

$$\rho_d^{1+\alpha}=\frac{B}{1+A}\left[\frac{A_s+(1-A_s)(1+z)^{3(1+\alpha)(1+A)}}{KA_s-(1-K)(1-A_s)(1+z)^{3(1+\alpha)(1+A)}}\right] \tag{19}$$

where A_s is constant satisfying $1-K<A_s<1$. So the present value of the energy density is $\rho_{CJ0}^{1+\alpha}=\frac{B}{(1+A)[KA_s-(1-K)(1-A_s)]}$.
If the DE of the Universe obeys the MCJG equation of state (1), then using the equation (19), the equation (12) reduces to the following equation for $E(z)$ as

$$\begin{aligned}E^2(z)=\Omega_{r0}(1+z)^4+\Omega_{m0}(1+z)^3-\Omega_{k0}(1+z)^2+\Omega_{CJ0}[KA_s-(1-K)(1-A_s)]^{\frac{1}{1+\alpha}}\\ \times\left[\frac{A_s+(1-A_s)(1+z)^{3(1+\alpha)(1+A)}}{KA_s-(1-K)(1-A_s)(1+z)^{3(1+\alpha)(1+A)}}\right]^{\frac{1}{1+\alpha}}-2f'(T)-\frac{f(T)}{6H_0^2}\end{aligned} \tag{20}$$

where we have defined the normalized Hubble parameter $E(z)=\frac{H(z)}{H_0}$ and the dimensionless density parameters $\Omega_{r0}=\frac{8\pi G\rho_{r0}}{3H_0^2}$, $\Omega_{m0}=\frac{8\pi G\rho_{m0}}{3H_0^2}$, $\Omega_{k0}=\frac{k}{H_0^2}$ and $\Omega_{CJ0}=\frac{8\pi G\rho_{CJ0}}{3H_0^2}$. Putting $z=0$ in the above equation, we obtain

$\Omega_{r0}+\Omega_{m0}-\Omega_{k0}+\Omega_{CJ0}-2f'(T_0)-\frac{f(T_0)}{6H_0^2}=1$ where $T_0=-6H_0^2\,(1+\Omega_{k0})$.

• **MCAG :** Next, we consider the DE in the form of modified Chaplygin-Abel gas (MCJG) whose equation of state is given by [71]

$$p_d=[(b^2+2c^2)(1+A)-1]\rho_d-\frac{c^2B}{\rho_d^{\alpha}}-\frac{(b^2+c^2)(1+A)^2}{B}\;\rho_d^{2+\alpha} \tag{21}$$

where $b,c\in\mathbb{R}$. Putting the above expression in (17), we obtain

$$\rho_d^{1+\alpha}=\frac{c^2B}{1+A}\left[\frac{B_s+(1-B_s)(1+z)^{3b^2(1+\alpha)(1+A)}}{c^2B_s+(b^2+c^2)(1-B_s)(1+z)^{3b^2(1+\alpha)(1+A)}}\right] \tag{22}$$

where B_s is a constant satisfying $0<B_s<1$. So the present value of the energy density is $\rho_{CA0}^{1+\alpha}=\frac{c^2B}{(1+A)[c^2B_s+(b^2+c^2)(1-B_s)]}$.

If the DE of the Universe obeys the MCAG equation of state (2), then using the equation (22), the equation (12) reduces to the following equation for $E(z)$ as

$$\begin{aligned}E^2(z)=\Omega_{r0}(1+z)^4+\Omega_{m0}(1+z)^3-\Omega_{k0}(1+z)^2+\Omega_{CA0}[c^2B_s+(b^2+c^2)(1-B_s)]^{\frac{1}{1+\alpha}}\\ \times\left[\frac{B_s+(1-B_s)(1+z)^{3b^2(1+\alpha)(1+A)}}{c^2B_s+(b^2+c^2)(1-B_s)(1+z)^{3b^2(1+\alpha)(1+A)}}\right]^{\frac{1}{1+\alpha}}-2f'(T)-\frac{f(T)}{6H_0^2}\end{aligned} \tag{23}$$

where the dimensionless density parameter $\Omega_{CA0}=\frac{8\pi G\rho_{CA0}}{3H_0^2}$. Putting $z=0$ in the above equation, we obtain $\Omega_{r0}+\Omega_{m0}-\Omega_{k0}+\Omega_{CA0}-2f'(T_0)-\frac{f(T_0)}{6H_0^2}=1$.

III. VIABLE $f(T)$ MODEL

Theoretically, the function $f(T)$ can be elected independently; but in practice it should be chosen in accordance with the fact the divergence from TEGR should be very small. Keeping that in mind, various feasible models were proposed over the years. Bengochea and Ferraro have introduced the power-law model characterized by [43]:

$$f(T)=\beta(-T)^{\gamma} \tag{24}$$

where β and γ are the two model parameters. For $\beta=0$, spatially flat Friedmann equation is recovered. To attain accelerated expansion from this model, the restriction $\gamma<1$ need to be imposed.

• **MCJG:** From equation (20), we obtain

$$\begin{aligned}E^2(z)+6\beta H_0^2\left[E^2(z)+\Omega_{k0}(1+z)^2-2\gamma\right]\left[E^2(z)+\Omega_{k0}(1+z)^2\right]^{\gamma-1}\\ =\Omega_{r0}(1+z)^4+\Omega_{m0}(1+z)^3-\Omega_{k0}(1+z)^2+\Omega_{CJ0}[KA_s-(1-K)(1-A_s)]^{\frac{1}{1+\alpha}}\\ \times\left[\frac{A_s+(1-A_s)(1+z)^{3(1+\alpha)(1+A)}}{KA_s-(1-K)(1-A_s)(1+z)^{3(1+\alpha)(1+A)}}\right]^{\frac{1}{1+\alpha}}\end{aligned} \tag{25}$$

with $\Omega_{r0}+\Omega_{m0}-\Omega_{k0}+\Omega_{CJ0}=1+6\beta H_0^2(1+\Omega_{k0}-2\gamma)(1+\Omega_{k0})^{\gamma-1}$.

• **MCAG:** From equation (23), we obtain

$$\begin{aligned}E^2(z)+6\beta H_0^2\left[E^2(z)+\Omega_{k0}(1+z)^2-2\gamma\right]\left[E^2(z)+\Omega_{k0}(1+z)^2\right]^{\gamma-1}\\ =\Omega_{r0}(1+z)^4+\Omega_{m0}(1+z)^3-\Omega_{k0}(1+z)^2+\Omega_{CA0}[c^2B_s+(b^2+c^2)(1-B_s)]^{\frac{1}{1+\alpha}}\\ \times\left[\frac{B_s+(1-B_s)(1+z)^{3b^2(1+\alpha)(1+A)}}{c^2B_s+(b^2+c^2)(1-B_s)(1+z)^{3b^2(1+\alpha)(1+A)}}\right]^{\frac{1}{1+\alpha}}\end{aligned} \tag{26}$$

with $\Omega_{r0} + \Omega_{m0} - \Omega_{k0} + \Omega_{CA0} = 1 + 6\beta H_0^2(1 + \Omega_{k0} - 2\gamma)(1 + \Omega_{k0})^{\gamma-1}$.

IV. DATA ANALYSIS

In this section, we will perform a comprehensive comparison between the predictions of the Modified Chaplygin-Abel gas (MCAG) and Modified Chaplygin-Jacobi gas (MCJG) models with observational data. Our objective is to understand the constraints placed on these models by utilizing different distinct datasets: the $H(z)$ dataset, the ype Ia supernova, Gamma Ray Bursts (GRBs), quasa (Q) , Baryon Acoustic Oscillation (BAO). By comparing the model predictions with these cosmological data sets, We aim to determine the best-fit values for free parameters of each model. In our analysis, we will also incorporate the present-day value of the Hubble function H_0 to provide a robust estimation of the model parameters. To achieve this, we adopt a rigorous and widely-used approach based on Bayesian statistics. This statistical technique involves the use of likelihood functions to quantify the agreement between the model predictions and the observational data. By applying the Markov Chain Monte Carlo (MCMC) method within the Bayesian framework, we can efficiently explore the parameter space and sample the probability distributions of each parameters. The Bayesian approach offers several advantages, including the ability to include prior information or constraints from previous studies and to handle uncertainties in the model and observational data. By employing this method, we can obtain robust constraints on the parameters of the MCAG and MCJG models and gain valuable insights into their viability in explaining the observed cosmological phenomena. The statistical analysis will allow us to determine the regions in parameter space where the models are consistent with the data and, consequently, provide valuable guidance for understanding the nature of dark energy and its role in the expansion of the Universe.

A. Methodology

Markov Chain Monte Carlo (MCMC) methods have become indispensable tools in cosmology for exploring parameter spaces and estimating uncertainties. In the field of cosmology, MCMC techniques offer a powerful approach to sampling from high-dimensional posterior distributions, enabling us to infer cosmological parameters and understand the underlying models. MCMC algorithms have proven particularly valuable for fitting theoretical models to observed data, such as cosmic microwave background measurements, large-scale structure surveys, and type Ia supernova data. By comparing model predictions to observational data, MCMC methods allow us to constrain parameter values and quantify their uncertainties. The MCMC process involves constructing a Markov chain, where each point represents a specific set of parameter values. These values are randomly generated based on a proposal distribution and accepted or rejected based on the likelihood of the data given the model and prior information. Through iterations, the Markov chain converges to the target posterior distribution, providing samples that represent parameter estimates. The strength of MCMC lies in its ability to efficiently explore complex parameter spaces, even in the presence of multimodal or degenerate likelihood surfaces. By generating a large number of samples from the posterior distribution, we can estimate parameter uncertainties, derive credible intervals, and assess model goodness of fit. In cosmology, MCMC techniques have been extensively employed to investigate various topics, including the properties of dark matter and dark energy, inflationary models, primordial power spectra, and cosmological parameters like the Hubble constant, matter density, and dark energy equation of state.

B. Data Discription

1. $H(z)$ measurements

In our analysis, we have utilized the Hubble expansion rate data to impose more stringent constraints on our Dark Energy (DE) models. The Hubble expansion rate, denoted by $H(z)$, describes how the Universe's expansion rate changes with cosmic time or redshift, and it is a critical parameter for understanding the evolution of the cosmos. We can derive $H(z)$ data through two main methods: the clustering of galaxies and quasars by measuring the Baryon Acoustic Oscillations (BAO) in the radial direction [72], or the differential age method, which expresses the Hubble parameter as follows:

$$H(z) = -\frac{1}{1+z}\frac{dz}{dt}, \tag{27}$$

where $\frac{dz}{dt}$ can be inferred from the redshift difference Δz and the age difference Δt between two passively evolving galaxies [73]. For the current analysis, we have employed a compilation of 36 data points of the Hubble parameter, which are presented in table II of the reference [74]. Each data point in the compilation is associated with its corresponding reference. To quantify the agreement between the model predictions and the observed values of $H(z)$, we utilize the $\chi^2_{H(z)}$ statistic, given by:

$$\chi^2_{H(z)} = \sum_{i=1}^{36} \frac{\left[H_{\rm th}\left(z_i\right) - H_{\rm obs}\left(z_i\right)\right]^2}{\sigma^2_{H(z_i)}}, \tag{28}$$

where $H_{\rm th}$ represents the model predictions, $H_{\rm obs}$ represents the measured values, and $\sigma_{H(z_i)}$ is the standard deviation associated with each data point at redshift z_i.

2. *The type Ia supernova dataset*

The study of Type Ia supernovae (SNIa) has played a crucial role in our understanding of cosmic accelerated expansion. SNIa has proven to be one of the most efficient methods for investigating the nature of dark energy, which remains one of the most intriguing and mysterious phenomena in cosmology. Over the past few years, a wealth of supernova data has been collected, providing valuable insights into the expansion of the Universe [75–80]. The recent release of a comprehensive dataset comprising 1701 SNIa observations is a significant milestone in this field. These observations span an extensive redshift range from $0.001 < z < 2.3.$, allowing researchers to explore the Universe's expansion history over a broad cosmic time scale. SNIa are exceptionally luminous astronomical events, and they are considered to be nearly perfect standard candles. This unique property makes them valuable tools for measuring relative distances using the distance modulus. The distance modulus relates the intrinsic brightness of the supernova to its observed apparent brightness, enabling scientists to calculate precise distances to distant objects and thus probe the expansion rate of the Universe at different epochs. The chi-square values of the SNIa dataset are given by

$$\chi^2_{Pantheon+} = \vec{D}^T \,.\, \mathbf{C}^{-1}_{Pantheon+} \,.\, \vec{D}, \tag{29}$$

where $\mathbf{C}_{Pantheon+}$ is the covariance matrix provided with the Pantheon+ data, including both statistical and systematic uncertainties. and

$$\vec{D} = m_{Bi} - M - \mu_{\rm model}, \tag{30}$$

where m_{Bi} and M denote the apparent, and the absolute magnitudes of SNIa, respectively. Furthermore, $\mu_{\rm model}$ denotes the corresponding distance modulus as predicted by an assumed cosmological model, and defined as

$$\mu_{\rm model}\left(z_i\right) = 5\log_{10}\frac{D_L(z_i)}{(H_0/c)Mpc} + 25, \tag{31}$$

where H_0 is the current value of the Hubble rate, and D_L is the luminosity distance, expressed, for a flat homogeneous and isotropic FLRW Universe, as follows

$$D_L(z) = (1+z)H_0\int_0^z \frac{dz'}{H\left(z'\right)}. \tag{32}$$

Unlike the Pantheon sample, the degeneracy between the absolute magnitude M and H_0 is broken in the Pantheon+ dataset, due to the rewriting of the vector $\vec{D}$, appearing in Eq. (29), in terms of the distance moduli of SNIa in the Cepheid hosts, which allows to constrain M independently via the relations

$$\vec{D}'_i = \begin{cases} m_{Bi} - M - \mu_i^{\rm Ceph} & i \in \text{ Cepheid hosts} \\ m_{Bi} - M - \mu_{\rm model}\ (z_i) & \text{otherwise }, \end{cases} \tag{33}$$

where $\mu_i^{\rm Ceph}$ stands for the distance modulus corresponding to the Cepheid host of the $i^{\rm th}$ SNIa, measured independently with Cepheid calibrators. Therefore, Eq. (29) can be rewritten as

$$\chi^2_{SN} = \vec{D'}^T \cdot \mathbf{C}^{-1}_{\rm Pantheon+} \cdot \vec{D'} \tag{34}$$

MCMC Results							
Model	Parameters	Priors	Bestfit Value	Model	Parameters	Priors	Bestfit Value
ΛCDM	H_0	[50.,100.]	$69.854848^{+1.259100}_{-1.259100}$	ΛCDM	H_0	[50.,100.]	$69.158620^{+0.354533}_{-0.454345}$
	$\Omega_{\rm m0}$	[0., 1.]	$0.268654^{+0.012822}_{-0.015435}$		$\Omega_{\rm m0}$	[50., 100.]	$0.268654^{+0.035225}_{-0.012822}$
MCJG	H_0	[50.,100.]	$69.365373^{+2.493749}_{-1.493749}$	MCAG	H_0	[50.,100.]	$69.158620^{+1.593749}_{-1.343566}$
	$\Omega_{\rm m0}$	[0.,1.]	$0.203134^{+0.0356345}_{-0.045126}$		$\Omega_{\rm m0}$	[0.,1.]	$0.322633^{+0.043225}_{-0.0545433}$
	$\Omega_{\rm r0}$	[0.,1.]	$0.293748^{+0.0034524}_{-0.001282}$		$\Omega_{\rm r0}$	[0.,1.]	$0.489371^{+0.0443255}_{-0.023543}$
	$\Omega_{\rm k0}$	[0.,1.]	$0.134637^{+0.004322}_{-0.001233}$		$\Omega_{\rm k0}$	[0.,1.]	$0.285599^{+0.045126}_{-0.0445635}$
	$\Omega_{\rm CJ0}$	[0.,1.]	$0.477354^{+0.001787}_{-0.003242}$		$\Omega_{\rm CAO}$	[0.,1.]	$0.700436^{+0.06754}_{-0.07655}$
	A	[-10.,-9.]	$-9.234245^{+0.009433}_{-0.0084643}$		A	[-8.,-7.]	$-7.543525^{+0.044255}_{-0.045324}$
	α	[1.5,2.5]	$2.093362^{+0.009647}_{-0.008374}$		α	[1.5,2.5]	$2.142141^{+0.065435}_{-0.044232}$
	β	[-0.1,0.1]	$-0.025066^{+0.043245}_{-0.053542}$		B_s	[2.,3.]	$2.534432^{+0.069934}_{-0.0155433}$
	γ	[0.,1.]	$0.567318^{+0.095311}_{-0.086122}$		b	[1.,2.]	$1.481343^{+0.067483}_{-0.094873}$
	A_s	[0.5,1.5]	$0.959363^{+0.006647}_{-0.009647}$		c	[0.,1.]	$0.694735^{+0.093853}_{-0.067843}$
	K	[0.,1.]	$0.692843^{+0.006432}_{-0.003532}$		β	[0.,0.5]	$0.0377415^{+0.07584}_{-0.067785}$
					γ	[0.,1.]	$0.6427995^{+0.069383}_{-0.079584}$

We also have involve 162 Gamma Ray Bursts (GRBs) ranging in the redshift from 1.44 to 8.1 [81]. GRBs are incredibly bright and short-lived events, making them observable even from the farthest reaches of the visible Universe. In addition to the GRBs, We also included 24 observations of compact radio quasars [82]. These quasars were found in the redshift range of 0.46 to 2.76 and were identified through a survey of 613 ultra-compact radio sources at GHz frequencies using Very Long Baseline Interferometry (VLBI). Both GRBs and compact radio quasars are essential objects in cosmology. GRBs, being among the most luminous events in the Universe, provide valuable information about the distant regions of the cosmos. On the other hand, compact radio quasars offer insights into the properties of active galactic nuclei and their evolution over cosmic time. The combination of these datasets, along with other cosmological probes like Type Ia supernovae, allows us to gain a deeper understanding of the expansion history and the nature of dark energy.

3. Baryon Acoustic Oscillations (BAO)

We have used 17 BAO data points (please see table 1 of this work [83]) measures from [83]) measures from [84–95] a whole catalog of BAO (333) measurements have been selected because considering the whole catalog of BAO might result in significant inaccuracy because of data correlations; hence, a representative subset to minimize errors. Transverse BAO experiments produce measurements. of $D_H(z)/r_d = c/H(z)r_d$ with a comoving angular diameter distance.[96, 97].

$$D_M = \frac{c}{H_0} S_k \left(\int_0^z \frac{dz'}{E\left(z'\right)} \right), \tag{35}$$

where

$$S_k(x) = \begin{cases} \frac{1}{\sqrt{\Omega_k}} \sinh\left(\sqrt{\Omega_k} x\right) & \text{if} \quad \Omega_k > 0 \\ x & \text{if} \quad \Omega_k = 0 \\ \frac{1}{\sqrt{-\Omega_k}} \sin\left(\sqrt{-\Omega_k} x\right) & \text{if} \quad \Omega_k < 0. \end{cases} \tag{36}$$

We also consider the angular diameter distance $D_A = D_M/(1+z)$ and the $D_V(z)/r_d$. This corresponds to the combination of the BAO peaked coordinates and the sound horizon r_d at the drag epoch. Furthermore, we could immediately derive "line-of-sight" (or "radial") observations from the Hubble parameter.

$$D_V(z) \equiv \left[z D_H(z) D_M^2(z) \right]^{1/3}. \tag{37}$$

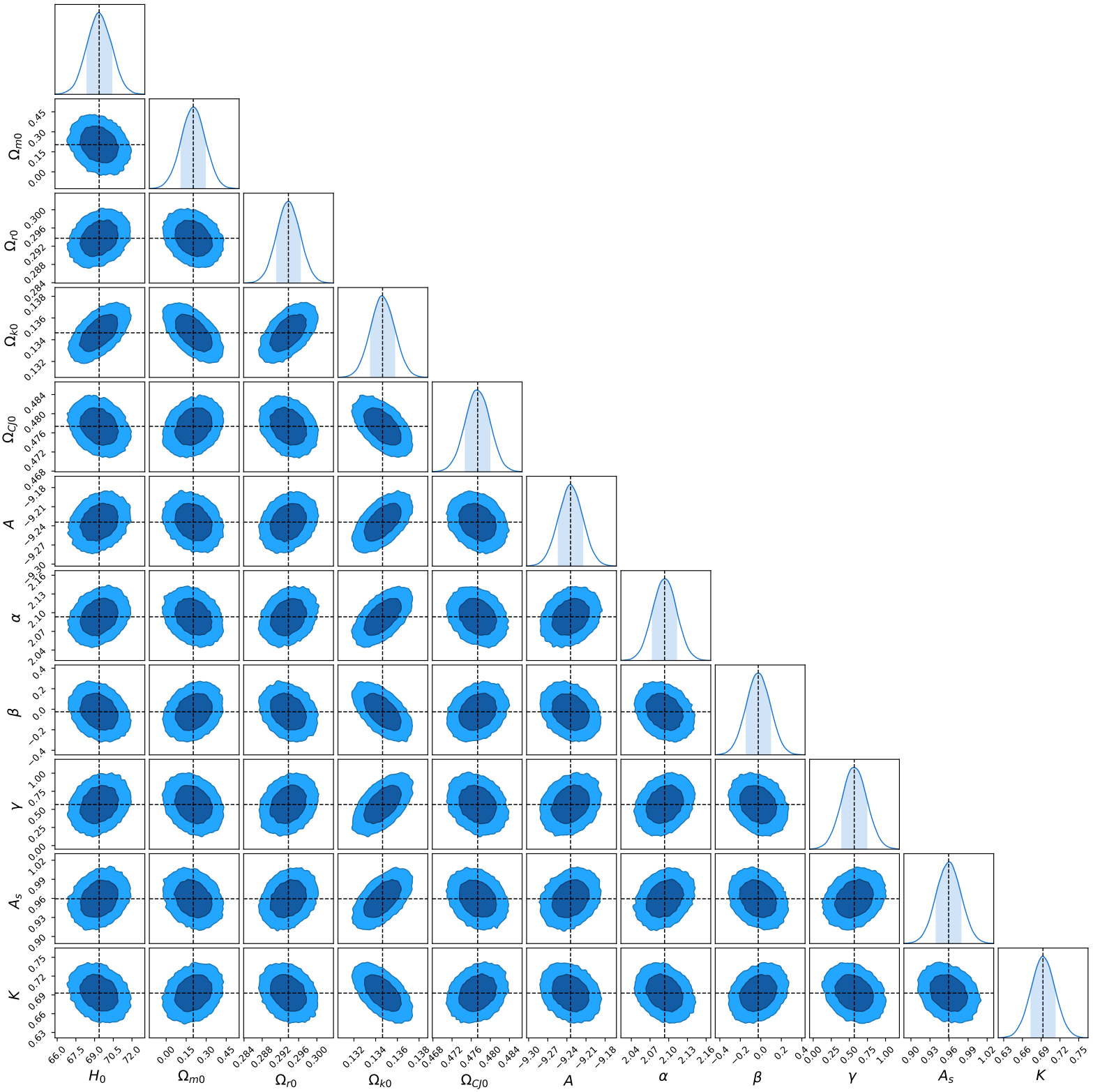


FIG. 1. The above figure shows the MCMC confidence contours at 1σ and 2σ of MCJG Model

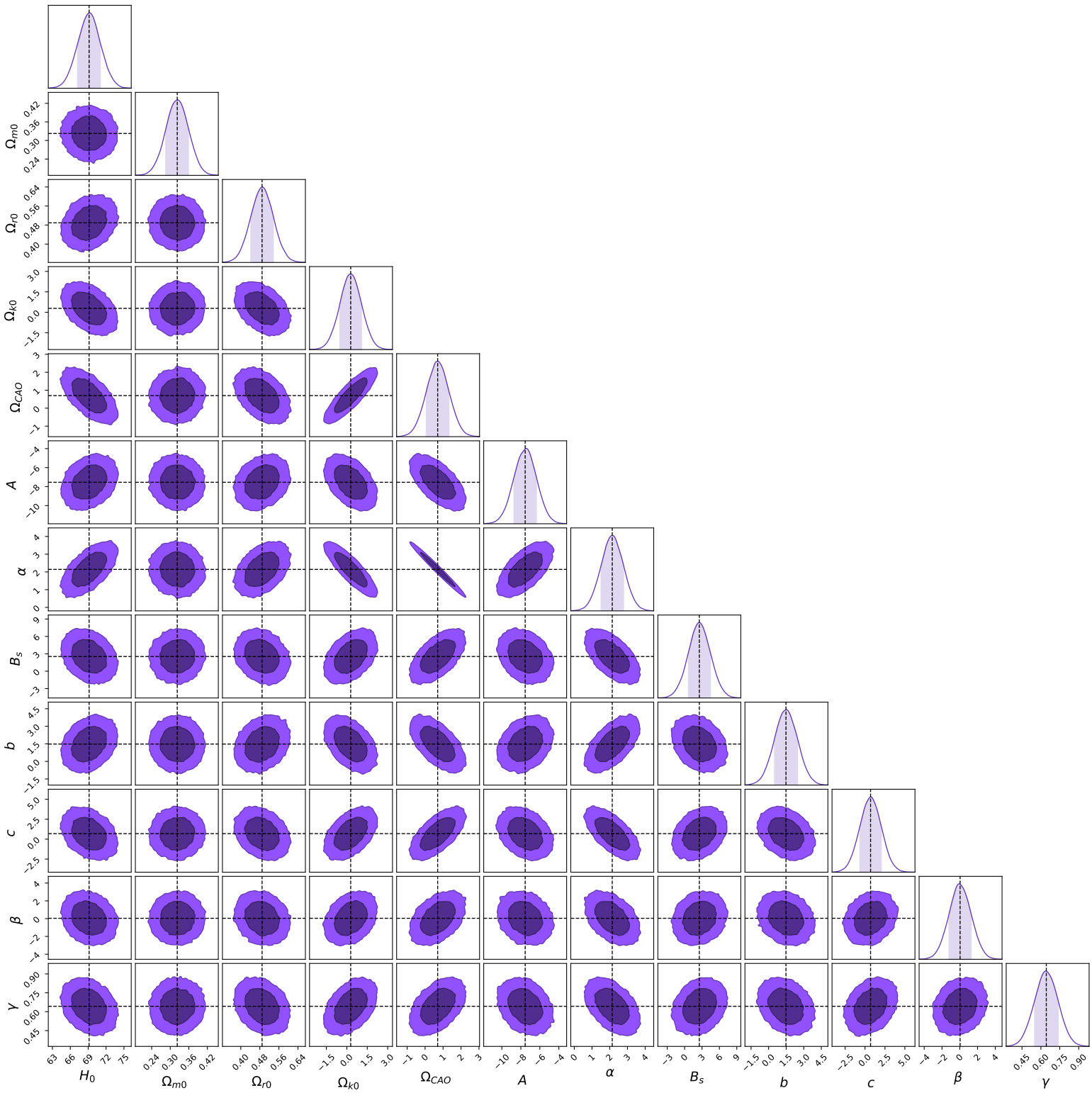


FIG. 2. The above figure shows the MCMC confidence contours at 1σ and 2σ of MCAG Model

V. OBSERVATIONAL AND THEORETICAL COMPARISONS OF THE HUBBLE FUNCTION AND DISTANCE MODULUS FUNCTION

After obtaining the best-fit values for the model parameters of the MCJG and MCAG Models, it is crucial to compare these models with the widely accepted ΛCDM model, which has demonstrated remarkable consistency with various observational datasets. The ΛCDM model is considered a robust framework for describing the evolution of the Universe and serves as a reference point for assessing the viability and reliability of alternative models. By comparing MCJG and MCAG Models with the ΛCDM model, we can investigate the deviations and discrepancies between the two and gain insights into the unique features and behaviors of MCJG and MCAG Models. This comparative analysis allows us to examine how MCJG and MCAG Models differ from the ΛCDM model in terms of the expansion rate, matter content, and overall dynamics of the Universe. Analyzing the deviations between MCJG and MCAG Models and the ΛCDM model provides valuable information about the strengths and limitations of MCJG and MCAG Models. It helps us understand the specific aspects of MCJG and MCAG Models that introduce variations compared to the well-established ΛCDM framework. By identifying these deviations, we can gain insights into the potential implications of MCJG and MCAG Models for our understanding of the Universe and explore novel cosmological dynamics and characteristics. Furthermore, the comparison with the ΛCDM model allows us to assess the goodness-of-fit of MCJG and MCAG Models to observational data. By evaluating the level of agreement between MCJG and MCAG Models models and the ΛCDM framework, we can determine the validity and reliability of MCJG and MCAG Models in light of the available observational constraints.

A. Comparison of MCJG and MCAG with Hubble Measurements

To evaluate the agreement between MCJG and MCAG models with observational data, we compared their predictions to the Hubble data, alongside the well-established ΛCDM model. The results are presented in Fig:- 3 and Fig:- 4. The figures clearly show that both MCJG and MCAG models fit the Hubble data quite well. The data points obtained from observations match very well with the predictions of our models, indicating that both models accurately explain how the Universe is expanding. The fact that both MCJG and MCAG models match the Hubble data provides strong support for their viability and indicates that they capture essential aspects of cosmic expansion. These findings demonstrate the satisfactory agreement of both models with the Hubble data, underscoring their potential to provide meaningful insights into the dynamics and evolution of the Universe within the framework of our proposed models. Such a positive outcome reinforces the credibility and significance of our research and its ability to contribute to a better understanding of the Universe's evolution.

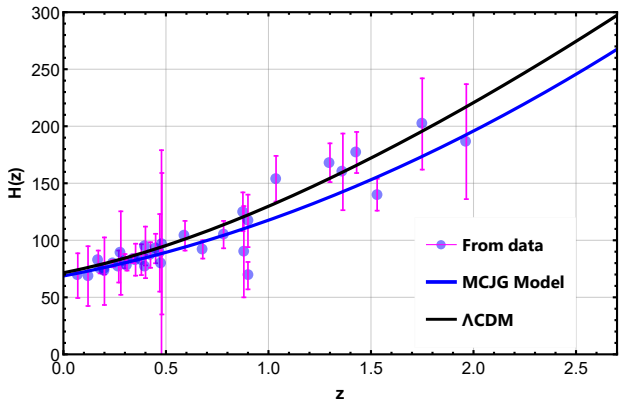


FIG. 3. The figure shows that the theoretical curve of the Hubble function $H(z)$ of MCJG Model shown in blue line and ΛCDM model shown in black line with $\Omega_{m0} = 0.3$ and $\Omega_\Lambda = 0.7$ against the 36 Hubble measurements are shown in blue dots wit their corresponding error bars in purple line.

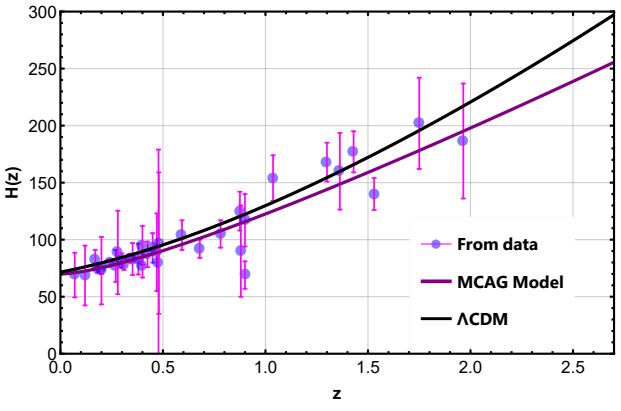


FIG. 4. The figure shows that the theoretical curve of the Hubble function $H(z)$ of MCAG Model shown in purple line and ΛCDM model shown in black line with $\Omega_{m0} = 0.3$ and $\Omega_\Lambda = 0.7$, against 36 $H(z)$ datasetsthe 36 Hubble measurements are shown in blue dots wit their corresponding error bars in purple line.

B. Comparison of MCJG and MCAG with type Ia supernova

In this analysis, we have compared two distance modulus functions, labeled as MCJG and MCAG, with the type Ia supernova dataset, which consists of 1701 data points, as well as the standard ΛCDM model. The comparison results are illustrated in Fig:- 5 and Fig:- 6. These figures shows that both MCJG and MCAG provide an excellent fit to the type Ia supernova dataset and the ΛCDM model. This means that our models closely match the observed distance measurements, and they are consistent with the data we have collected. This comparison with real observational data provides strong support for the validity and trustworthiness of our models in explaining the phenomena we observe in the Universe. It gives us confidence that both MCJG and MCAG accurately represent the expansion history of the Universe, as indicated by the type Ia supernova dataset.

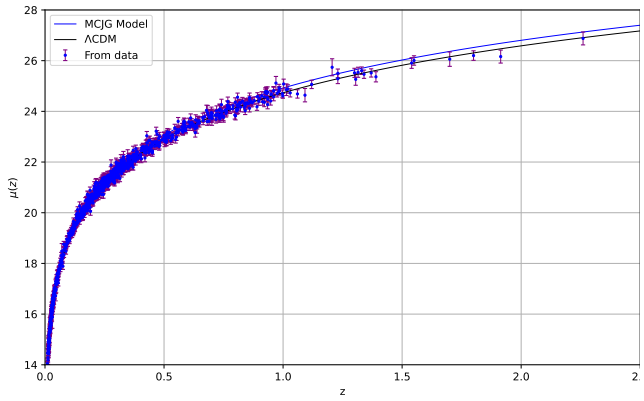


FIG. 5. Theoretical curve of distance modulus $\mu(z)$ of the MCJG is shown in blue line and the ΛCDM model is shown in the black line with $\Omega_{m0} = 0.3$ and $\Omega_\Lambda = 0.7$, against type Ia supernova data are shown in blue dots with their corresponding errors bars shown in purple.

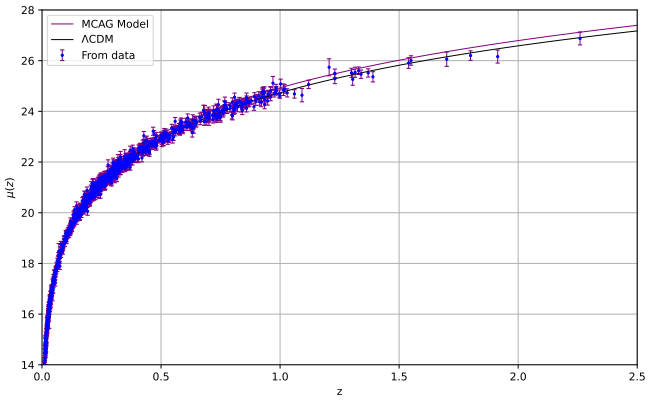


FIG. 6. Theoretical curve of distance modulus $\mu(z)$ of the MCAG is shown in purple line and the ΛCDM model is shown in the black line with $\Omega_{m0} = 0.3$ and $\Omega_\Lambda = 0.7$, against type Ia supernova data are shown in blue dots with their corresponding errors bars shown in purple.

C. Relative difference between MCJG and MCAG with ΛCDM

We have examined the relative difference between MCAJG and MCAG, and the standard ΛCDM paradigm. These comparisons are visualized in Fig:- 7 and Fig:- 8. The figures reveal interesting insights regarding the performance of the typical ΛCDM model and the two alternative models for different redshifts. Specifically, for redshifts z less than 1, all three models demonstrate very similar behavior, indicating a close agreement in their predictions. However, as we move to higher redshifts ($z > 1$), discrepancies begin to emerge between the two alternative models and the ΛCDM model. The differences become more pronounced as the redshift increases. This suggests that at higher redshifts, the alternative models deviate from the predictions of the ΛCDM model.

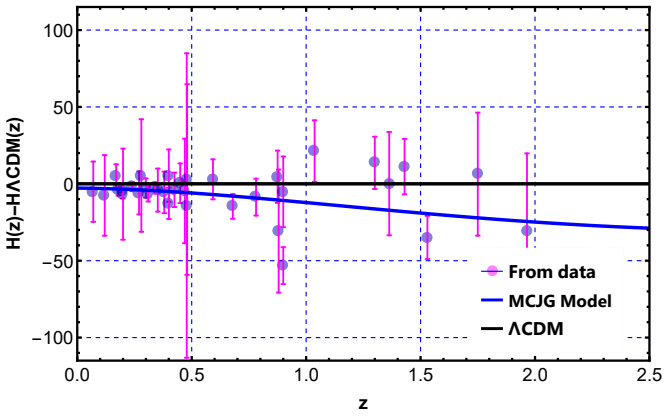


FIG. 7. The variation of the difference between MCJG shown in the blue line, and the ΛCDM model shown in the black line with $\Omega_{m0} = 0.3$ and $\Omega_\Lambda = 0.7$, as a function of the redshift z against the 36 Hubble measurements are shown in blue dots with their corresponding error bars shown in purple line.

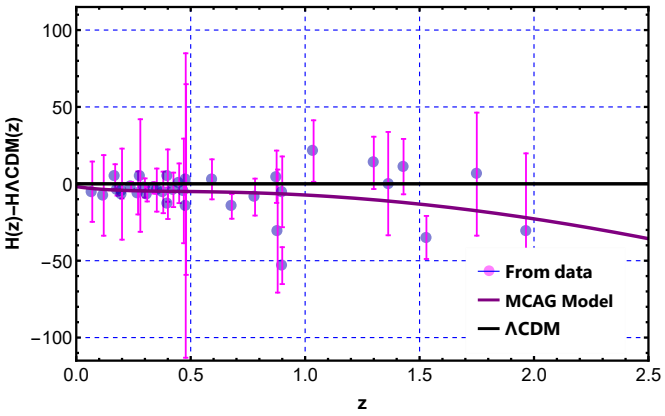


FIG. 8. The variation of the difference between MCAG shown in the purple line, and the ΛCDM model shown in the black line with $\Omega_{m0} = 0.3$ and $\Omega_\Lambda = 0.7$, as a function of the redshift z against the 36 Hubble measurements are shown in blue dots with their corresponding error bars shown in purple line.

VI. GEOMETRICAL DYNAMICS OF THE MODELS

A. Cosmographic analysis

Cosmography parameters [98–100] provide a valuable framework for analyzing and comparing theoretical cosmological models with observational data. They offer a comprehensive set of quantities that describe the expansion history of the Universe and can be used to test the consistency of different cosmological models. The cosmography approach is based on a Taylor expansion of the scale factor, which represents the size of the Universe, around a reference redshift or time. By expanding the scale factor up to a certain order, we can obtain a series of cosmography parameters that capture important aspects of the cosmic evolution. These parameters are derived from observables such as the Hubble parameter and its derivatives. Some of the commonly used cosmography parameters include the Hubble parameter ($H(z)$), deceleration parameter ($q(z)$), jerk parameter ($j(z)$), and snap parameter ($s(z)$), among others. These parameters can be derived from the expansion history of the Universe and provide a quantitative description of its behavior. They allow us to extract meaningful information about the underlying cosmological model and compare it with observational data.

1. *deceleration parameter*

The deceleration parameter is a fundamental cosmological quantity that describes the rate at which the expansion of the Universe is slowing down or speeding up. It provides valuable insights into the dynamics of the cosmic expansion and plays a crucial role in understanding the past and future behavior of the Universe. In the context of Friedmann-Robertson-Walker cosmology, the deceleration parameter, denoted as $q(z)$, is defined as the negative ratio of the second derivative of the scale factor $a(t)$ to its first derivative, normalized by the square of the Hubble parameter $H(z)$:

$$q(z) = -\frac{a''(t)}{a(t)H(t)^2} \tag{38}$$

Here, $a(t)$ represents the scale factor of the Universe, which describes how the distances between galaxies change with time, and $H(t)$ is the Hubble parameter, which characterizes the expansion rate of the Universe at a given time. The deceleration parameter $q(z)$ determines the behavior of the cosmic expansion. If $q(z) < 0$, the Universe is experiencing accelerated expansion, indicating that the expansion rate is increasing over time. Conversely, if $q(z) > 0$, the Universe is undergoing decelerated expansion, with the expansion rate decreasing over time. Studying the deceleration parameter is essential for understanding the evolution of the Universe and the influence of different components such as matter, radiation, or dark energy on its expansion dynamics. References such as [101–105] have

employed the deceleration parameter to analyze and explain the evolutionary phases of the Universe in different cosmological models.

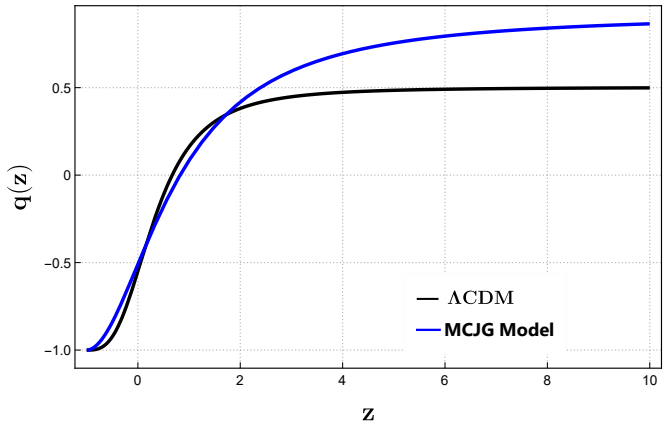


FIG. 9. Evolution of deceleration parameter with respect to the redshift of MCJG Model.

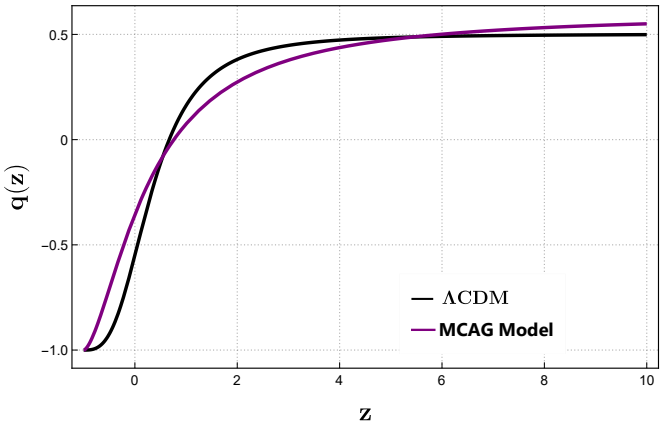


FIG. 10. Evolution of deceleration parameter with respect to the redshift of MCAG Model.

2. Jerk parameter

The concept of jerk parameters in cosmology provides valuable insights into the dynamics of the Universe and the nature of dark energy. Jerk parameters offer novel diagnostics for dark energy, allowing us to study the higher derivatives of the scale factor and explore deviations from the standard cosmological model. In cosmology, the dimensionless jerk parameter, sometimes referred to as jolt, extends the standard cosmological parameters, such as the scale factor $a(t)$ and the deceleration parameter q. It arises from the fourth term of a Taylor series expansion of the scale factor around a reference time t_0. Mathematically, it can be expressed as:

$$\frac{a(t)}{a_0} = 1 + H_0(t-t_0) - \frac{1}{2}q_0H_0^2(t-t_0)^2 + \frac{1}{6}j_0H_0^3(t-t_0)^3 + O\left[(t-t_0)^4\right], \tag{39}$$

where the subscript 0 denotes the present values. The jerk parameter j characterizes the third-order derivative of the expansion factor with respect to cosmic time and offers an absolute measure to explore departures from the standard ΛCDM model. It can be expressed as:

$$j = \frac{1}{a}\frac{d^3a}{d\tau^3}\left[\frac{1}{a}\frac{da}{d\tau}\right]^{-3} = q(2q+1) + (1+z)\frac{dq}{dz}, \tag{40}$$

where τ represents the cosmic time and z is the redshift. This expression provides a direct way to determine the jerk parameter by relating it to the deceleration parameter q and its derivative with respect to redshift. The jerk parameter is a valuable diagnostic tool in cosmology as it helps determine the most suitable candidate for the physical interpretation of cosmic dynamics, particularly in relation to dark energy models. Different values of the jerk parameter offer insights into the correlation between various dark energy proposals and standard Universe models. In the flat ΛCDM model, the jerk parameter takes the value $j = 1$. Deviations from this value indicate departures from the standard model and provide a way to explore transitions between different phases of cosmic acceleration.

3. snap parameter

The snap parameter is a measure of the fifth time derivative of the expansion factor in cosmology. It is a dimensionless quantity that provides information about the curvature of the Universe and the dynamics of its expansion. The snap parameter is associated with the term in the Taylor expansion of the expansion factor beyond the third order. In the context of the Taylor expansion of the expansion factor, the snap parameter represents the fourth-order

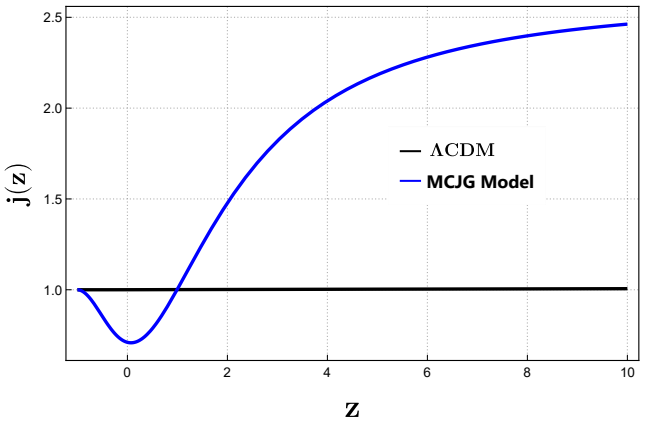


FIG. 11. Evolution of jerk parameter with respect to the redshift of MCJG Model.

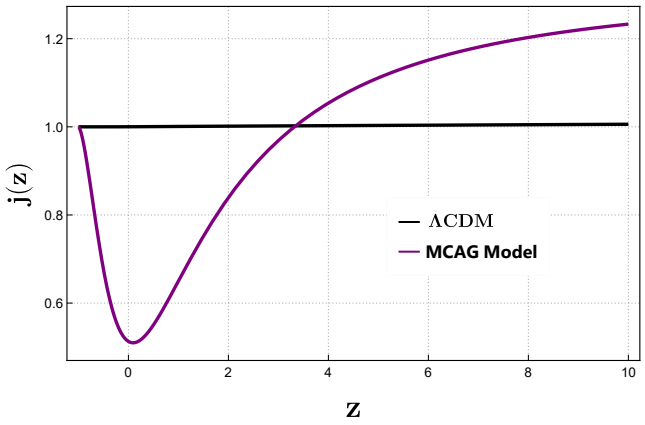


FIG. 12. Evolution of jerk parameter with respect to the redshift of MCAG Model.

term. The expansion factor is expressed as a series expansion around a reference time t_0, with the snap parameter denoted as s_0. The expansion factor can be written as:

$$\frac{a(t)}{a_0} = 1 + H_0(t-t_0) - \frac{1}{2}q_0H_0^2(t-t_0)^2 + \frac{1}{6}j_0H_0^3(t-t_0)^3 + \frac{1}{24}s_0H_0^4(t-t_0)^4 + O\left[(t-t_0)^5\right] \tag{41}$$

Here, $a(t)$ represents the expansion factor at time t, a_0 is the expansion factor at the reference time t_0, H_0 is the present-day Hubble constant, q_0 is the deceleration parameter, j_0 is the jounce parameter (related to the snap parameter), and s_0 is the snap parameter. The higher-order terms beyond the fourth order are neglected in this expansion. The snap parameter s can also be expressed in terms of the deceleration parameter q and the jounce parameter j as:

$$s = \frac{1}{a}\frac{d^4a}{d\tau^4}\left[\frac{1}{a}\frac{da}{d\tau}\right]^{-4} = \frac{j-1}{3\left(q-\frac{1}{2}\right)} \tag{42}$$

For the flat ΛCDM model, where $j = 1$, the snap parameter simplifies to $s = -(2+3q)$. In this case, the deviation of the quantity $\frac{ds}{dq}$ from -3 quantifies how the cosmic evolution diverges from the dynamics predicted by the ΛCDM model.

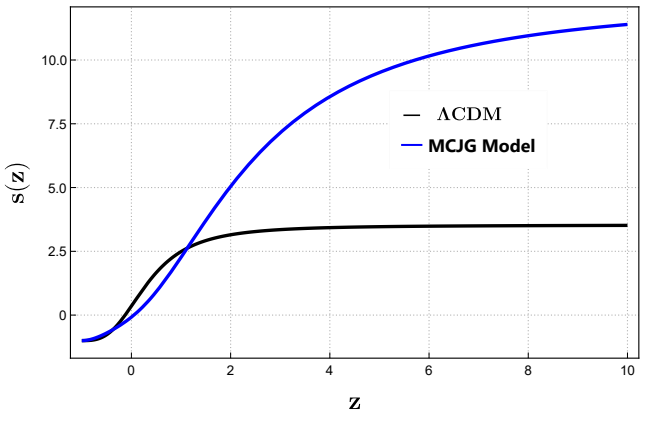


FIG. 13. Evolution of snap parameter with respect to the redshift of MCJG Model.

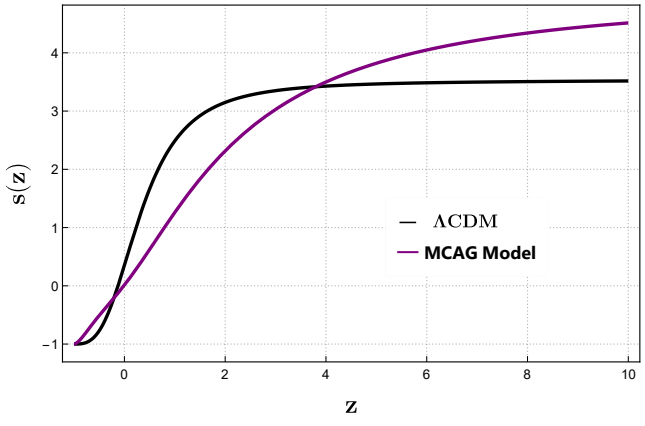


FIG. 14. Evolution of snap parameter with respect to the redshift of MCAG Model.

VII. STATEFINDER DIAGNOSTIC

The statefinder diagnostic pair, consisting of the parameters r and s, is a widely-used tool in cosmology for studying and understanding various dark energy (DE) models. It provides a dimensionless way to analyze the cosmic properties of DE, independent of specific models, by examining higher-order derivatives of the scale factor. This diagnostic pair was introduced in the literature ([106–109]). The parameter r in the statefinder diagnostic represents the ratio of the third derivative of the scale factor to the cube of the Hubble parameter. It quantifies the acceleration dynamics of the Universe. On the other hand, the parameter s is a linear combination of r and the deceleration parameter q. It characterizes the cosmic evolution and the nature of DE. The statefinder diagnostic pair can be calculated using the following expressions:

$$r = \frac{\dddot{a}}{aH^3}, \quad s = \frac{r-1}{3\left(q-\frac{1}{2}\right)}, \tag{43}$$

Here, H is the Hubble parameter, and q is the deceleration parameter. By computing the values of r and s, one can analyze the cosmic characteristics of DE models without relying on specific theoretical frameworks. Certain values of the statefinder pair have well-known interpretations in the context of standard DE models. For example, $\{r, s\} = \{1, 0\}$ corresponds to the ΛCDM model, which is the standard cosmological model with a cosmological constant. Similarly, $\{r, s\} = \{1, 1\}$ corresponds to the standard cold dark matter (SCDM) model in a Friedmann-Lemaître-Robertson-Walker (FLRW) Universe. The region with $\{r, s\}$ values ranging from $(-\infty, \infty)$ corresponds to the Einstein static Universe. In the r-s plane, different regions represent distinct types of DE models. For example, when $s > 0$, it indicates quintessence-like models, which involve a scalar field responsible for DE. Conversely, when $s < 0$, it indicates phantom-like models, characterized by the presence of a scalar field with negative kinetic energy. Deviations from the standard $\{r, s\} = \{1, 0\}$ values can signify an evolutionary process from phantom-like to quintessence-like behavior. Moreover, specific combinations of the deceleration parameter q and the statefinder parameter r are associated with well-known models. For instance, $\{q, r\} = \{-1, 1\}$ is linked to the ΛCDM model, which describes a Universe dominated by dark energy in the form of a cosmological constant. Another example is $\{q, r\} = \{0.5, 1\}$, corresponding to the SCDM model, which assumes a Universe dominated by non-relativistic matter without dark energy.

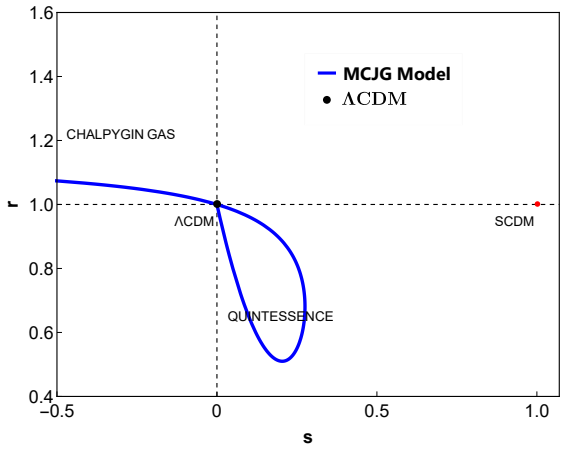


FIG. 15. This figure shows $\{s, r\}$ plots for MCJG Model .

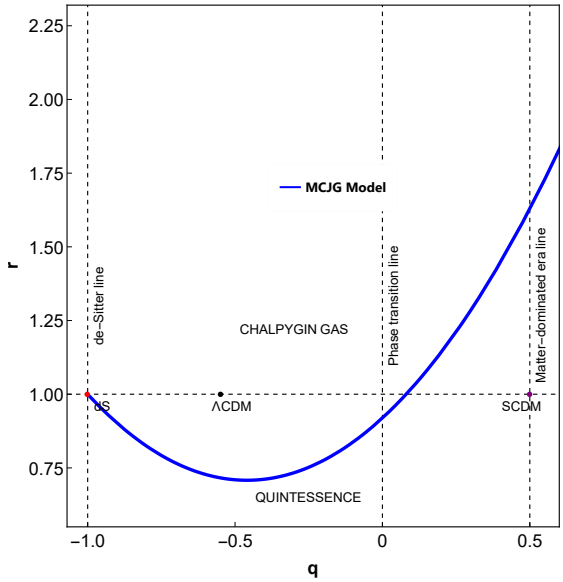


FIG. 16. This figure shows $\{q, r\}$ plots for MCJG Model.

VIII. DIAGNOSTIC ANALYSIS

The Om diagnostic is a valuable tool introduced in the literature ([110–113]) that serves as an effective null test for the ΛCDM (Lambda Cold Dark Matter) model. This diagnostic, similar to the statefinder diagnostic, offers a

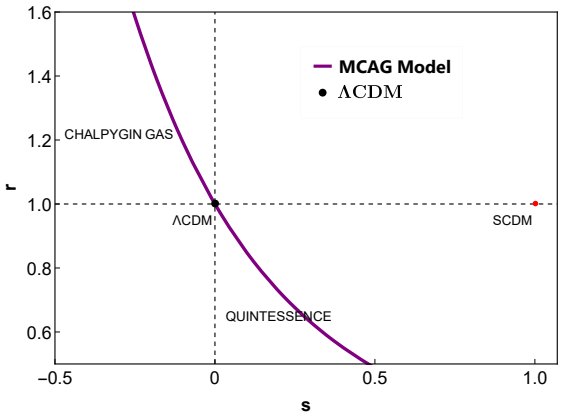


FIG. 17. This figure shows $\{s, r\}$ plots for MCAG Model .

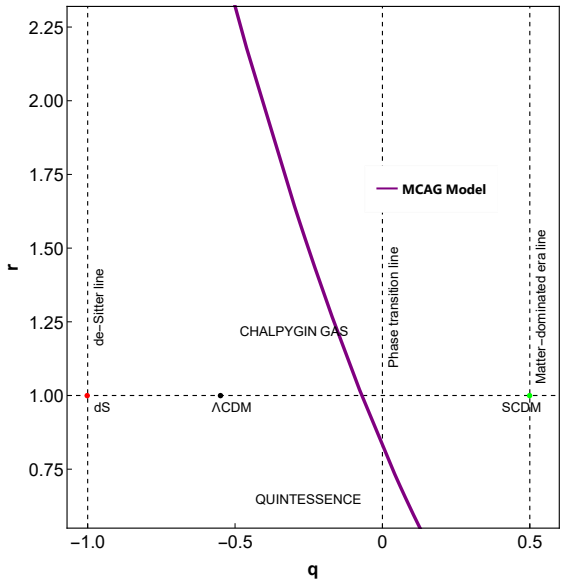


FIG. 18. This figure shows $\{q, r\}$ plots for MCAG Model.

viable approach for differentiating various dark energy (DE) models from the standard ΛCDM model by examining the slope variation of Om. The Om diagnostic provides insights into the nature of dark energy by characterizing its behavior and distinguishing between different regimes. The diagnostic is based on the analysis of the $\mathrm{Om}(z)$ function, where z represents the redshift, in a flat Universe. By studying the variation of $\mathrm{Om}(z)$ with respect to redshift, distinct features and properties of the underlying dark energy model can be identified. A negative slope in the Om diagnostic corresponds to a phantom nature of dark energy. This suggests that the equation of state of dark energy is less than -1, indicating an accelerated expansion with even stronger dynamics compared to the ΛCDM model. Conversely, a positive slope in the Om diagnostic implies a quintessence regime for dark energy. In this case, the equation of state is greater than -1, indicating a slower rate of expansion compared to the phantom scenario. Furthermore, a constant slope of the diagnostic curve with respect to redshift corresponds to the ΛCDM model. This indicates that the dark energy component behaves like a cosmological constant throughout cosmic history, providing a good fit to observational data. The Om diagnostic thus enables us to discern different dark energy models from the ΛCDM model based on the characteristic slope variation of Om. By analyzing observational data and comparing the slope of $\mathrm{Om}(z)$, researchers can gain valuable insights into the nature of dark energy and the validity of alternative models.

$$\mathrm{Om}(z) = \frac{\left(\frac{H(z)}{H_0}\right)^2 - 1}{(1+z)^3 - 1}. \tag{44}$$

IX. INFORMATION CRITERIA

To assess the viability of any model, a thorough understanding of information criteria (IC) is essential. The Akaike Information Criterion (AIC) [114–118] is commonly employed as a general IC tool. The AIC serves as an approximate measure of the Kullback-Leibler information divergence and asymptotically provides an unbiased estimator of this divergence. The AIC for Gaussian estimation is given by $\mathrm{AIC} = -2\ln(\mathcal{L}_{\max}) + 2\kappa + \frac{2\kappa(\kappa+1)}{N-\kappa-1}$, where $\mathcal{L}_{\max}$ denotes the maximum likelihood function, κ represents the total number of free parameters in the model, and N is the total number of data points used. Since the assumption is often $N \gg 1$ for the models considered, the formula simplifies to the original AIC form, $\mathrm{AIC} = -2\ln(\mathcal{L}_{\max}) + 2\kappa$. When comparing multiple models, the differences in IC values can be quantified as $\triangle\mathrm{AIC} = \mathrm{AIC}_{\mathrm{model}} - \mathrm{AIC}_{\min} = \triangle\chi^2_{\min} + 2\triangle\kappa$. The range of $\triangle$AIC that is considered more favorable is $(0, 2)$. The lower favorable range of $\triangle$AIC is $(4, 7)$, while values exceeding $\triangle\mathrm{AIC} > 10$ provide less support for the

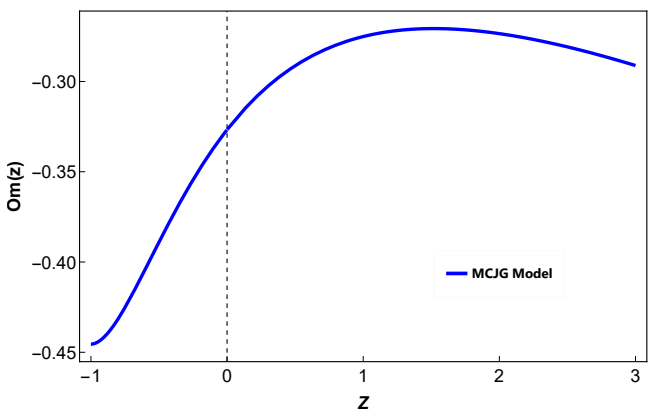


FIG. 19. This figure shows the $Om(z)$ with respect to redshift for MCJG Model.

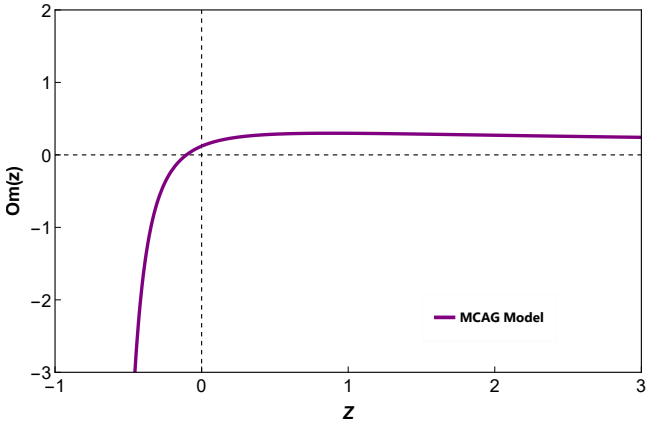


FIG. 20. This figure shows the $Om(z)$ with respect to redshift for MCAG Model.

model.

Model	χ^2_{min}	χ^2_{red}	AIC	ΔAIC
ΛCDM Model	2357.562	0.9837	1507.84	0
MCJG Model	2339.187	0.9827	1507.66	0.8202
MCAG Model	2338.167	0.9813	1507.81	1.9741

TABLE I. Summary of the χ^2_{min}, χ^2_{red}, AIC and ΔAIC.

X. RESULTS

a. Deceleration parameter The comparison of the redshift dependence of the deceleration parameter for MCJG Model and MCJG Model with respect to the ΛCDM model is presented in Fig:- 10 and Fig:- 9, respectively. In the considered redshift range of $z \in (-1, 10)$, all three models exhibit similar behavior in the evolution of the deceleration parameter. However, in the MCJG Model , noticeable deviations from the ΛCDM model can be observed at high redshifts beyond $z = 4$. This indicates that the MCJG Model may introduce differences in the expansion dynamics compared to the standard ΛCDM model. The numerical values of the transition redshift, denoted as z_{tr}, which marks the transition from a decelerating phase to an accelerating phase, are relatively close for all three models, including the ΛCDM model. This implies that, in terms of the deceleration parameter, the models share similar characteristics regarding the shift from deceleration to acceleration in cosmic expansion. Finally, it is noteworthy that all three models exhibit a de Sitter phase characterized by a deceleration parameter of $q = -1$. The de Sitter phase corresponds to a phase of accelerated expansion driven by a cosmological constant or a similar dark energy component. These findings suggest that while the MCJG Model and the MCAJ Model generally resemble the behavior of the ΛCDM model in terms of the deceleration parameter, deviations become apparent at high redshifts in the MCJG Model. However, all models share similarities in terms of the transition redshift z_{tr} and end in a de Sitter phase with $q = -1$, indicating accelerated expansion.

b. Jerk parameter The behavior of the jerk parameter, denoted as $j(z)$, for the (MCJG) Model and (MCAG) Model in comparison to the standard ΛCDM paradigm is depicted in Fig:- 11 and Fig:- 12, respectively. These models provide alternative frameworks to explain the dynamics of the Universe and explore departures from the ΛCDM model. In Fig:- 11, we observe that the predictions of the MCJG Model diverge significantly from the ΛCDM model at high redshifts. The MCJG Model exhibits a distinct behavior, showing a different trend in the evolution of cosmic acceleration. However, it is worth noting that at $z = -1$, both the MCJG Model and the ΛCDM model predict the same value for the jerk parameter. Similarly, in Fig:- 12, we examine the behavior of the jerk parameter for the MCAG Model. This model also presents deviations from the ΛCDM model, particularly at higher redshifts.

The MCAG Model showcases an opposite trend compared to the MCAG Model, suggesting a unique dynamic for cosmic acceleration. Remarkably, at $z = -1$, both the MCAG Model and the ΛCDM model yield identical values for the jerk parameter.

c. Snap parameter The behavior of the snap parameter, represented as $s(z)$, for the MCJG Model and MCAG Model, in comparison to the ΛCDM model, is depicted in Figure 13 and Figure 14, respectively. These figures clearly illustrate significant deviations between the alternative models and the ΛCDM model, particularly at higher redshifts. In the context of cosmology, the snap parameter, $s(z)$, characterizes the rate of change of the acceleration of the Universe with respect to redshift. It is an important quantity for discerning the dynamic nature of dark energy. Analyzing Figure 13, we observe that the MCJG Model exhibits discrepancies compared to the ΛCDM model, with the differences becoming less pronounced at lower redshifts. This suggests that the MCJG Model aligns more closely with the ΛCDM model as redshift decreases, indicating a better agreement in terms of the snap parameter. On the other hand, Figure 14 illustrates that the MCAG Model displays a similar behavior to the ΛCDM model. The snap parameter for the MCAG Model exhibits characteristics that are consistent with the ΛCDM model, implying a comparable agreement in terms of the dynamics of acceleration. These observations imply that the MCJG Model and MCAG Model have distinct behaviors in comparison to the ΛCDM model, as evidenced by the variations in the snap parameter. While the MCJG Model converges towards the ΛCDM model at lower redshifts, the MCAG Model demonstrates a level of agreement with the ΛCDM model throughout the redshift range.

d. Statefinder diagnostic In Figure 15, the evolution of the statefinder pair (r, s) for the MCJG Model is depicted. As the evolution progresses, the statefinder trajectory of the MCJG Model transitions from the quintessence region to the Chaplygin gas region at late times. This transition implies a change in the nature of dark energy, reflecting a shift in the dominant energy component driving the accelerated expansion of the Universe. It is worth noting that during this evolution, the MCJG Model bypasses the intermediate fixed point $(0, 1)$, which is associated with the standard ΛCDM model. Similarly, in Figure 17, the evolution of the statefinder pair (r, s) for the MCAG Model is presented. The behavior of the MCAG Model is analogous to the MCJG Model as it undergoes a transition from the quintessence region to the Chaplygin gas region. This transition is again accompanied by bypassing the intermediate fixed point $(0, 1)$, which characterizes the ΛCDM model. In the context of cosmology, the quintessence region corresponds to a scenario where dark energy behaves like a dynamical scalar field, while the Chaplygin gas region represents a different regime where dark energy is modeled as a perfect fluid with an equation of state given by the Chaplygin gas equation. The trajectory of the statefinder pair (r, s) provides valuable insights into the nature and evolution of dark energy within these alternative models. Fig 16 provides additional insight into the behavior of the MCJG Model. As the evolution progresses, the trajectory of the MCJG Model undergoes a transition from the line corresponding to the Matter-dominated era. This region is characterized by a decelerating expansion due to the dominant matter content in the Universe. As the MCJG Model evolves further, it enters the quintessence region, which is indicated by the deviation from the Matter-dominated era line. In this region, the dominant energy component responsible for the cosmic acceleration is a quintessence scalar field with an equation of state greater than -1. This scalar field drives the accelerated expansion of the Universe. The trajectory of the MCJG Model ultimately converges towards the de Sitter point, denoted as $(-1, 1)$. At this point, the expansion of the Universe becomes asymptotically exponential, mimicking a cosmological constant (dark energy with a constant equation of state equal to -1). On the other hand, Fig 18 illustrates the evolution of the statefinder pair (q, r) for the MCAG Model. The trajectory of the MCAG Model undergoes a transition from the quintessence region, similar to the MCJG Model. However, in the case of the MCAG Model, the trajectory deviates towards the de Sitter point $(-1, 1)$. This trajectory reveals the unique characteristics of the MCAG Model and its departure from the simple ΛCDM model.

e. Om diagnostic Figures 19 and 20 depict the behavior of the $\mathrm{Om}(z)$ parameter with respect to redshift z for the MCJG Model and MCAG Model, respectively. Analyzing Figure 19, we observe that for the MCJG Model, the $\mathrm{Om}(z)$ values are smaller than Ω_{m0} for redshifts greater than zero ($z > 0$). This indicates that the MCJG Model resides in the quintessence domain, where the expansion of the Universe is driven by a slowly-evolving quintessence field or dark energy component. As the redshift decreases, we notice a significant decrease in the $\mathrm{Om}(z)$ value for the MCJG Model. Eventually, the value becomes negative, signifying a transition from the quintessence regime to the phantom region. The phantom region is characterized by an equation of state for dark energy that is less than -1, leading to an accelerated expansion with more dramatic dynamics than the standard ΛCDM model. Similarly, in Figure 20 representing the MCAG Model, we observe that the $\mathrm{Om}(z)$ values are also smaller than Ω_{m0} for $z > 0$, placing the model within the quintessence domain. As the redshift decreases, the $\mathrm{Om}(z)$ value exhibits a significant

decrease and eventually becomes negative, indicating the transition from quintessence to the phantom region within the MCAG Model. The behavior observed in the transition from quintessence to phantom regions for both the MCJG and MCAG Models provides valuable insights into the cosmological dynamics and characteristics of these models. The change in the sign of Om(z) signifies a shift from a slowly-evolving quintessence field to a more rapidly-evolving dark energy component, leading to an accelerated expansion of the Universe with distinct dynamics compared to the standard ΛCDM model.

f. Information Criteria In our analysis, we compared the MCJG and MCAG models with the well-established ΛCDM model, and we assessed their relative support using the ΔAIC values. Table I presents the results, showing that MCJG has a ΔAIC value of 0.8202, while MCAG has a ΔAIC value of 1.9741. Positive ΔAIC values indicate that both MCJG and MCAG have slightly weaker support compared to the ΛCDM model. When the ΔAIC value falls within the range of (0, 2), it is considered most favorable, suggesting moderate evidence in favor of a model. In our case, both MCJG and MCAG fall within this range, indicating some degree of support for these models. Comparing the ΛCDM model to MCJG, we find that MCJG has a higher ΔAIC value, indicating that the ΛCDM model performs slightly better in terms of goodness of fit and model selection. Similarly, when comparing the ΛCDM model to MCAG, MCAG has a higher ΔAIC value, suggesting a better fit for the ΛCDM model. Our analysis suggests that the ΛCDM model performs slightly better than MCJG and MCAG based on the ΔAIC values.

XI. CONCLUSION

In this article, we consider a non-flat FLRW model of the Universe filled with radiation, dark matter and dark energy in the context of $f(T)$ gravity, which was proposed based on the corrections of the teleparallel equivalent of General Relativity (TEGR). It's 2nd order equations of motion makes it simpler than $f(R)$ gravity. Here we choose Modified Chaplygin-Jacobi gas (MCJG) and modified Chaplygin-Abel gas (MCAG) to play the role of dark energy. Also here we choose the power law form as viable $f(T)$ gravity model amongst various feasible $f(T)$ gravity models proposed in literature as this form exhibits both matter dominated and radiation dominated era and successfully confronts the cosmological test with observational data. Here we obtain the Constraints on the model parameters for DE as well as the $f(T)$ gravity models by fitting recent astronomical data. Exploiting the observational data coming from H(z), Pantheon, Gamma Ray Bursts, quasars and BAO measurements, we have determined the best fit values of the model parameters. We present the reduced Hubble parameter in terms of observable parameters like $\Omega_{r0}, \Omega_{m0}, \Omega_{k0}, \Omega_{CJ0}, \Omega_{CA0}, H_0$ for both the considered dark energy models. In the figures 1 and 2, we have drawn the contour plots of the MCMC confidence contours at 1σ and 2σ of MCJG model and MCAG model respectively. The best fit values of the parameters of MCJG model and MCAG model have been obtained by χ^2 minimum test as (i) H_0 = 69.158620, $\Omega_{\rm m0}$ = 0.190482, $\Omega_{\rm k0}$ = 0.001891, $\Omega_{\rm cjo}$ = 0.001891, α = 1.000928 β = 1.000928, A_s = -0.929042, K = 0.402158, γ=1.000928. (ii) H_0 = 69.158620, $\Omega_{\rm m0}$ = 0.190482 ,$\Omega_{\rm k0}$ = 0.001891, $\Omega_{\rm cjo}$ = 0.001891, α=1.000928, β=1.000928, A_s=-0.929042, K=0.402158, γ=1.000928. Contrasting the predicted values of the considered cosmological models with the Hubble 57 observational dataset and ΛCDM paradim, we have seen the models adequetly matches the Hubble and the Pantheon dataset. In conclusion, our analysis of the MCJG Model and MCAG Model in comparison to the ΛCDM model has provided valuable insights into the dynamics of cosmic expansion and the nature of dark energy. Through various diagnostic tools, including the deceleration parameter, jerk parameter, snap parameter, and statefinder diagnostic, we have examined the behavior of these alternative models and their deviations from the standard model. The comparison of the deceleration parameter revealed that while all models share similarities in terms of the transition from deceleration to acceleration, noticeable differences were observed at high redshifts in the MCJG Model. This suggests that the MCJG Model introduces variations in the expansion dynamics compared to the ΛCDM model. However, all models ultimately converge to a de Sitter phase with a deceleration parameter of $q = -1$, indicating accelerated expansion driven by a cosmological constant or a similar dark energy component. The analysis of the jerk parameter showcased distinct behaviors of the MCJG Model and MCAG Model compared to the ΛCDM model, particularly at higher redshifts. These models exhibited different trends in cosmic acceleration, highlighting their departure from the standard model. However, at $z = -1$, both alternative models aligned with the ΛCDM model, suggesting similarities in the acceleration dynamics at that specific redshift. The examination of the snap parameter emphasized the differences between the MCJG Model and the ΛCDM model, with a closer agreement observed at lower redshifts. In contrast, the MCAG Model displayed behavior consistent with the ΛCDM model throughout the entire redshift range. These findings underscore the unique characteristics of these models in terms of their acceleration dynamics. Furthermore, the statefinder diagnostic provided insights into the nature of dark energy within the MCJG Model and MCAG Model. The transitions from the quintessence region to the Chaplygin gas region indicated changes in the dominant energy component responsible for cosmic acceleration. These transitions

were accompanied by deviations from the intermediate fixed point associated with the ΛCDM model. The statefinder trajectories shed light on the evolution and behavior of dark energy within these alternative models. Our analysis highlights the distinct behaviors and deviations of the MCJG Model and MCAG Model from the ΛCDM model. These alternative models offer valuable insights into the dynamics of cosmic expansion and the properties of dark energy. While the MCJG Model exhibited notable differences at high redshifts, both models ultimately converged to a de Sitter phase characterized by accelerated expansion. The comparison of MCJG and MCAG with the ΛCDM model using ΔAIC values indicates moderate support for both models. While the ΛCDM model shows slightly better performance, further evidence from other observations is crucial to fully assess the viability of MCJG and MCAG in explaining dark energy and the universe's expansion. Our study has contributed to the ongoing quest for a deeper understanding of dark energy and its potential role in the cosmic acceleration. While the ΛCDM model remains a robust and successful framework for explaining various cosmological observations, exploring alternative models like MCJG and MCAG is vital for advancing our comprehension of the universe's fundamental properties. Our results open avenues for future research and inspire further investigations into the intriguing phenomena shaping the evolution of our cosmos

[1] S. Perlmutter, et al., Discovery of a supernova explosion at half the age of the Universe and its cosmological implications, Nature 391 (1998) 51–54. arXiv:astro-ph/9712212, doi:10.1038/34124.

[2] S. Perlmutter, et al., Measurements of Ω and Λ from 42 high redshift supernovae, Astrophys. J. 517 (1999) 565–586. arXiv:astro-ph/9812133, doi:10.1086/307221.

[3] A. G. Riess, et al., Observational evidence from supernovae for an accelerating universe and a cosmological constant, Astron. J. 116 (1998) 1009–1038. arXiv:astro-ph/9805201, doi:10.1086/300499.

[4] D. J. Eisenstein, W. Hu, Baryonic features in the matter transfer function, Astrophys. J. 496 (1998) 605. arXiv:astro-ph/9709112, doi:10.1086/305424.

[5] D. J. Eisenstein, et al., Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies, Astrophys. J. 633 (2005) 560–574. arXiv:astro-ph/0501171, doi:10.1086/466512.

[6] S. L. Bridle, O. Lahav, J. P. Ostriker, P. J. Steinhardt, Precision cosmology? Not just yet..., Science 299 (2003) 1532. arXiv:astro-ph/0303180, doi:10.1126/science.1082158.

[7] D. N. Spergel, et al., First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters, Astrophys. J. Suppl. 148 (2003) 175–194. arXiv:astro-ph/0302209, doi:10.1086/377226.

[8] E. Komatsu, et al., Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation, Astrophys. J. Suppl. 180 (2009) 330–376. arXiv:0803.0547, doi:10.1088/0067-0049/180/2/330.

[9] P. J. E. Peebles, B. Ratra, The Cosmological Constant and Dark Energy, Rev. Mod. Phys. 75 (2003) 559–606. arXiv:astro-ph/0207347, doi:10.1103/RevModPhys.75.559.

[10] T. Padmanabhan, Cosmological constant—the weight of the vacuum, Physics Reports 380 (5-6) (2003) 235–320.

[11] S. Weinberg, The Cosmological Constant Problem, Rev. Mod. Phys. 61 (1989) 1–23. doi:10.1103/RevModPhys.61.1.

[12] P. J. Steinhardt, L. Wang, I. Zlatev, Cosmological tracking solutions, Physical Review D 59 (12) (1999) 123504.

[13] E. J. Copeland, M. Sami, S. Tsujikawa, Dynamics of dark energy, Int. J. Mod. Phys. D 15 (2006) 1753–1936. arXiv:hep-th/0603057, doi:10.1142/S021827180600942X.

[14] L. Amendola, S. Tsujikawa, Dark energy: theory and observations, Cambridge University Press, 2010.

[15] S. A. Chaplygin, Gas jets, no. 1063, National Advisory Committee for Aeronautics, 1944.

[16] H. S. et al., The end of unified dark matter?, Phys. Rev. D 69 (2004) 123524.

[17] R. B. et al., Are Chaplygin gases serious contenders for the dark energy??, Phys. Rev. D 68 (2003) 023515.

[18] A. S. M.C. Bento, O. Bertolami, Generalized Chaplygin Gas, Accelerated Expansion and Dark Energy-Matter Unification, Phys. Rev. D 66 (2002) 043507.

[19] H. Benaoum, Accelerated universe from modified chaplygin gas and tachyonic fluid, Universe 8 (7) (2022) 340.

[20] U. Debnath, A. Banerjee, S. Chakraborty, Role of modified chaplygin gas in accelerated universe, Classical and Quantum Gravity 21 (23) (2004) 5609.

[21] Z.-K. Guo, Y.-Z. Zhang, Cosmology with a variable chaplygin gas, Physics Letters B 645 (4) (2007) 326–329.

[22] U. Debnath, Variable Modified Chaplygin Gas and Accelerating Universe, Astrophys. Space Sci. 312 (2007) 295–299. arXiv:0710.1708, doi:10.1007/s10509-007-9690-6.

[23] U. Debnath, M. Sayani, Correspondence of F-essence with Chaplygin gas cosmology, Eur. Phys. J. Plus 129 (2014) 14.

[24] X.-H. Zhai, Y.-D. Xu, X.-Z. Li, Viscous generalized Chaplygin gas, Int. J. Mod. Phys. D 15 (2006) 1151–1162. arXiv:astro-ph/0511814, doi:10.1142/S0218271806008784.

[25] C. Ranjit, S. Chakraborty, U. Debnath, Observational constraints of homogeneous higher-dimensional cosmology with modified chaplygin gas, The European Physical Journal Plus 128 (2013) 1–8.

[26] B. Paul, P. Thakur, Observational constraints on modified chaplygin gas from cosmic growth, Journal of Cosmology and Astroparticle Physics 2013 (11) (2013) 052.

[27] J. Bhadra, U. Debnath, Constraining the Parameters of New Variable Modified Chaplygin Gas Model, Int. J. Theor. Phys. 53 (2014) 1821–1831. doi:10.1007/s10773-013-1982-8.

[28] B. C. Paul, P. Thakur, A. Beesham, Constraints on modified Chaplygin gas from large scale structure, Astrophys. Space Sci. 361 (10) (2016) 336. arXiv:1410.6588, doi:10.1007/s10509-016-2916-8.

[29] D. Panigrahi, B. C. Paul, S. Chatterjee, Constraining Modified Chaplygin gas parameters, Grav. Cosmol. 21 (2015) 83. arXiv:1305.7204, doi:10.1134/S0202289315010119.

[30] N. Mebarki, D. Aberkane, Dynamical study of viscous modified chaplygin gas and confrontation with recent observational data, in: Journal of Physics: Conference Series, Vol. 1258, IOP Publishing, 2019, p. 012025.

[31] J. Villanueva, The generalized chaplygin-jacobi gas, Journal of Cosmology and Astroparticle Physics 2015 (07) (2015) 045.

[32] J. Villanueva, E. Gallo, A jacobian elliptic single-field inflation, The European Physical Journal C 75 (2015) 1–7.

[33] A. G. Cadavid, J. Villanueva, No slow-roll inflation à la generalized chaplygin gas in general relativity, Journal of Cosmology and Astroparticle Physics 2020 (12) (2020) 044.

[34] U. Debnath, Roles of modified Chaplygin–Jacobi and Chaplygin–Abel gases in FRW universe, Int. J. Mod. Phys. A 36 (33) (2021) 2150245. arXiv:2112.00296, doi:10.1142/S0217751X21502456.

[35] D. U. Mukherjee Puja, P. Anirudh, Accretion of Modified Chaplygin-Jacobi Gas and Modified Chaplygin-Abel Gas onto Schwarzschild Black Hole, Int. J. Geometric Methods in Mod. Phys. (2023). doi:10.1142/S0219887823502183.

[36] S. Capozziello, M. De Laurentis, Extended Theories of Gravity, Phys. Rept. 509 (2011) 167–321. arXiv:1108.6266, doi:10.1016/j.physrep.2011.09.003.

[37] S. Nojiri, S. D. Odintsov, Unified cosmic history in modified gravity: from F(R) theory to Lorentz non-invariant models, Phys. Rept. 505 (2011) 59–144. arXiv:1011.0544, doi:10.1016/j.physrep.2011.04.001.

[38] A. De Felice, S. Tsujikawa, f(R) theories, Living Rev. Rel. 13 (2010) 3. arXiv:1002.4928, doi:10.12942/lrr-2010-3.

[39] A. De Felice, S. Tsujikawa, Construction of cosmologically viable f(G) dark energy models, Phys. Lett. B 675 (2009) 1–8. arXiv:0810.5712, doi:10.1016/j.physletb.2009.03.060.

[40] P. Hořava, Membranes at quantum criticality, Journal of High Energy Physics 2009 (03) (2009) 020.

[41] D. Lovelock, The Einstein tensor and its generalizations, J. Math. Phys. 12 (1971) 498–501. doi:10.1063/1.1665613.

[42] R. Ferraro, F. Fiorini, On Born-Infeld Gravity in Weitzenbock spacetime, Phys. Rev. D 78 (2008) 124019. arXiv:0812.1981, doi:10.1103/PhysRevD.78.124019.

[43] G. R. Bengochea, R. Ferraro, Dark torsion as the cosmic speed-up, Phys. Rev. D 79 (2009) 124019. arXiv:0812.1205, doi:10.1103/PhysRevD.79.124019.

[44] E. V. Linder, Einstein's Other Gravity and the Acceleration of the Universe, Phys. Rev. D 81 (2010) 127301, [Erratum: Phys.Rev.D 82, 109902 (2010)]. arXiv:1005.3039, doi:10.1103/PhysRevD.81.127301.

[45] A. Unzicker, T. Case, Translation of Einstein's attempt of a unified field theory with teleparallelism (3 2005). arXiv:physics/0503046.

[46] C. Møller, Conservation laws and absolute parallelism in general relativity, Nordita Publ. 64 (1961) 50.

[47] C. Pellegrini, J. Plebanski, Tetrad fields and gravitational fields, Kgl. Danske Videnskab. Selskab, Mat. Fys. Skrifter 2 (4) (1963).

[48] K. Hayashi, T. Shirafuji, New General Relativity, Phys. Rev. D 19 (1979) 3524–3553, [Addendum: Phys.Rev.D 24, 3312–3314 (1982)]. doi:10.1103/PhysRevD.19.3524.

[49] R. Aldrovandi, J. G. Pereira, Teleparallel gravity: an introduction, Vol. 173, Springer Science & Business Media, 2012.

[50] J. W. Maluf, The teleparallel equivalent of general relativity, Annalen der Physik 525 (5) (2013) 339–357.

[51] R. Ferraro, F. Fiorini, Modified teleparallel gravity: Inflation without inflaton, Phys. Rev. D 75 (2007) 084031. arXiv:gr-qc/0610067, doi:10.1103/PhysRevD.75.084031.

[52] K. Bamba, S. Nojiri, S. D. Odintsov, Trace-anomaly driven inflation in $f(T)$ gravity and in minimal massive bigravity, Phys. Lett. B 731 (2014) 257–264. arXiv:1401.7378, doi:10.1016/j.physletb.2014.02.041.

[53] K. Bamba, G. G. L. Nashed, W. El Hanafy, S. K. Ibraheem, Bounce inflation in $f(T)$ Cosmology: A unified inflaton-quintessence field, Phys. Rev. D 94 (8) (2016) 083513. arXiv:1604.07604, doi:10.1103/PhysRevD.94.083513.

[54] K. Bamba, S. D. Odintsov, E. N. Saridakis, Inflationary cosmology in unimodular f (t) gravity, Modern Physics Letters A 32 (21) (2017) 1750114.

[55] S.-H. Chen, J. B. Dent, S. Dutta, E. N. Saridakis, Cosmological perturbations in f(T) gravity, Phys. Rev. D 83 (2011) 023508. arXiv:1008.1250, doi:10.1103/PhysRevD.83.023508.

[56] P. Wu, H. W. Yu, The dynamical behavior of $f(T)$ theory, Phys. Lett. B 692 (2010) 176–179. arXiv:1007.2348, doi:10.1016/j.physletb.2010.07.038.

[57] J. B. Dent, S. Dutta, E. N. Saridakis, f (t) gravity mimicking dynamical dark energy. background and perturbation analysis, Journal of Cosmology and Astroparticle Physics 2011 (01) (2011) 009.

[58] S. Basilakos, S. Nesseris, F. K. Anagnostopoulos, E. N. Saridakis, Updated constraints on f (t) models using direct and indirect measurements of the hubble parameter, Journal of Cosmology and Astroparticle Physics 2018 (08) (2018) 008.

[59] F. K. Anagnostopoulos, S. Basilakos, E. N. Saridakis, Bayesian analysis of $f(T)$ gravity using $f\sigma_8$ data, Phys. Rev. D 100 (8) (2019) 083517. arXiv:1907.07533, doi:10.1103/PhysRevD.100.083517.

[60] L. Iorio, E. N. Saridakis, Solar system constraints on f(T) gravity, Mon. Not. Roy. Astron. Soc. 427 (2012) 1555. arXiv:1203.5781, doi:10.1111/j.1365-2966.2012.21995.x.

[61] M. Jamil, U. Debnath, Interacting modified Chaplygin gas in loop quantum cosmology, Astrophys. Space Sci. 333 (2011) 3–8. arXiv:1102.2758, doi:10.1007/s10509-011-0651-8.

[62] M.-I. Park, A test of hořava gravity: the dark energy, Journal of Cosmology and Astroparticle Physics 2010 (01) (2010) 001.

[63] U. Debnath, Observational Constraints of Modified Chaplygin Gas in Chern-Simons Gravity, Int. J. Theor. Phys. 54 (1) (2015) 22–35. doi:10.1007/s10773-014-2195-5.

[64] C. Ranjit, P. Rudra, U. Debnath, Study of Some Parameters of Modified Chaplygin Gas in Galileon Gravity Theory from Observational Perspective, Can. J. Phys. 92 (12) (2014) 1667–1675. arXiv:1406.3043, doi:10.1139/cjp-2014-0287.

[65] U. Debnath, Constraining the Parameters of Modified Chaplygin Gas in Einstein-Aether Gravity, Adv. High Energy Phys. 2014 (2014) 653630. arXiv:1310.2144, doi:10.1155/2014/653630.

[66] U. Debnath, Constraining the parameters of modified chaplygin gas in brans–dicke theory, Physics of the Dark Universe 31 (2021) 100764.

[67] Y.-F. Cai, S.-H. Chen, J. B. Dent, S. Dutta, E. N. Saridakis, Matter Bounce Cosmology with the f(T) Gravity, Class. Quant. Grav. 28 (2011) 215011. arXiv:1104.4349, doi:10.1088/0264-9381/28/21/215011.

[68] M. Setare, F. Darabi, Power-law solutions in f (t) gravity, General Relativity and Gravitation 44 (2012) 2521–2527.

[69] S. Capozziello, O. Luongo, R. Pincak, A. Ravanpak, Cosmic acceleration in non-flat f (t) cosmology, General Relativity and Gravitation 50 (5) (2018) 53.

[70] U. Debnath, Constructions of entropy and modified friedmann equations in gravity theories, International Journal of Geometric Methods in Modern Physics 19 (06) (2022) 2250093.

[71] U. Debnath, Roles of modified chaplygin–jacobi and chaplygin–abel gases in frw universe, International Journal of Modern Physics A 36 (33) (2021) 2150245.

[72] E. Gaztanaga, A. Cabre, L. Hui, Clustering of luminous red galaxies–iv. baryon acoustic peak in the line-of-sight direction and a direct measurement of h (z), Monthly Notices of the Royal Astronomical Society 399 (3) (2009) 1663–1680.

[73] R. Jimenez, A. Loeb, Constraining cosmological parameters based on relative galaxy ages, The Astrophysical Journal 573 (1) (2002) 37.

[74] A. Bouali, I. Albarran, M. Bouhmadi-López, T. Ouali, Cosmological constraints of phantom dark energy models, Physics of the Dark Universe 26 (2019) 100391.

[75] M. Kowalski, D. Rubin, G. Aldering, R. Agostinho, A. Amadon, R. Amanullah, C. Balland, K. Barbary, G. Blanc, P. Challis, et al., Improved cosmological constraints from new, old, and combined supernova data sets, The Astrophysical Journal 686 (2) (2008) 749.

[76] R. Amanullah, C. Lidman, D. Rubin, G. Aldering, P. Astier, K. Barbary, M. Burns, A. Conley, K. Dawson, S. Deustua, et al., Spectra and hubble space telescope light curves of six type ia supernovae at 0.511¡ z¡ 1.12 and the union2 compilation, The Astrophysical Journal 716 (1) (2010) 712.

[77] N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, et al., The hubble space telescope cluster supernova survey. v. improving the dark-energy constraints above z¿ 1 and building an early-type-hosted supernova sample, The Astrophysical Journal 746 (1) (2012) 85.

[78] M. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin, R. Biswas, P. Astier, P. El-Hage, M. Konig, S. Kuhlmann, et al., Improved cosmological constraints from a joint analysis of the sdss-ii and snls supernova samples, Astronomy & Astrophysics 568 (2014) A22.

[79] D. M. Scolnic, D. Jones, A. Rest, Y. Pan, R. Chornock, R. Foley, M. Huber, R. Kessler, G. Narayan, A. Riess, et al., The complete light-curve sample of spectroscopically confirmed sne ia from pan-starrs1 and cosmological constraints from the combined pantheon sample, The Astrophysical Journal 859 (2) (2018) 101.

[80] D. Scolnic, D. Brout, A. Carr, A. G. Riess, T. M. Davis, A. Dwomoh, D. O. Jones, N. Ali, P. Charvu, R. Chen, et al., The pantheon+ type ia supernova sample: the full dataset and light-curve release, arXiv preprint arXiv:2112.03863 (2021).

[81] M. Demianski, E. Piedipalumbo, D. Sawant, L. Amati, Cosmology with gamma-ray bursts-i. the hubble diagram through the calibrated ep, i–eiso correlation, Astronomy & Astrophysics 598 (2017) A112.

[82] C. Roberts, K. Horne, A. O. Hodson, A. D. Leggat, Tests of \ λ cdm and conformal gravity using grb and quasars as standard candles out to z \ sim 8, arXiv preprint arXiv:1711.10369 (2017).

[83] D. Benisty, D. Staicova, Testing late-time cosmic acceleration with uncorrelated baryon acoustic oscillation dataset, Astronomy & Astrophysics 647 (2021) A38.

[84] W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Z. Ivezi'c, G. R. Knapp, et al., Baryon acoustic oscillations in the sloan digital sky survey data release 7 galaxy sample, Monthly Notices of the Royal Astronomical Society 401 (4) (2010) 2148–2168.

[85] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, F. Watson, The 6df galaxy survey: baryon acoustic oscillations and the local hubble constant, Monthly Notices of the Royal Astronomical Society 416 (4) (2011) 3017–3032.

[86] T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, D. Kirkby, J.-M. Le Goff, M. M. Pieri, A. Slosar, É. Aubourg, J. E. Bautista, et al., Baryon acoustic oscillations in the lyα forest of boss quasars, Astronomy & Astrophysics 552 (2013) A96.

[87] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden, A. J. Cuesta, et al., The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample, Monthly Notices of the Royal Astronomical Society 427 (4) (2012) 3435–3467.

[88] H.-J. Seo, S. Ho, M. White, A. J. Cuesta, A. J. Ross, S. Saito, B. Reid, N. Padmanabhan, W. J. Percival, R. De Putter, et al., Acoustic scale from the angular power spectra of sdss-iii dr8 photometric luminous galaxies, The Astrophysical Journal 761 (1) (2012) 13.

[89] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, M. Manera, The clustering of the sdss dr7 main galaxy sample–i. a 4 per cent distance measure at z= 0.15, Monthly Notices of the Royal Astronomical Society 449 (1) (2015)

835–847.
[90] R. Tojeiro, A. J. Ross, A. Burden, L. Samushia, M. Manera, W. J. Percival, F. Beutler, J. Brinkmann, J. R. Brownstein, A. J. Cuesta, et al., The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: galaxy clustering measurements in the low-redshift sample of data release 11, Monthly Notices of the Royal Astronomical Society 440 (3) (2014) 2222–2237.
[91] J. E. Bautista, M. Vargas-Magaña, K. S. Dawson, W. J. Percival, J. Brinkmann, J. Brownstein, B. Camacho, J. Comparat, H. Gil-Marín, E.-M. Mueller, et al., The sdss-iv extended baryon oscillation spectroscopic survey: baryon acoustic oscillations at redshift of 0.72 with the dr14 luminous red galaxy sample, The Astrophysical Journal 863 (1) (2018) 110.
[92] E. De Carvalho, A. Bernui, G. Carvalho, C. Novaes, H. Xavier, Angular baryon acoustic oscillation measure at z= 2.225 from the sdss quasar survey, Journal of Cosmology and Astroparticle Physics 2018 (04) (2018) 064.
[93] M. Ata, F. Baumgarten, J. Bautista, F. Beutler, D. Bizyaev, M. R. Blanton, J. A. Blazek, A. S. Bolton, J. Brinkmann, J. R. Brownstein, et al., The clustering of the sdss-iv extended baryon oscillation spectroscopic survey dr14 quasar sample: first measurement of baryon acoustic oscillations between redshift 0.8 and 2.2, Monthly Notices of the Royal Astronomical Society 473 (4) (2018) 4773–4794.
[94] T. Abbott, F. Abdalla, A. Alarcon, S. Allam, F. Andrade-Oliveira, J. Annis, S. Avila, M. Banerji, N. Banik, K. Bechtol, et al., Dark energy survey year 1 results: Measurement of the baryon acoustic oscillation scale in the distribution of galaxies to redshift 1, Monthly Notices of the Royal Astronomical Society 483 (4) (2019) 4866–4883.
[95] Z. Molavi, A. Khodam-Mohammadi, Observational tests of gauss-bonnet like dark energy model, The European Physical Journal Plus 134 (6) (2019) 254.
[96] N. B. Hogg, M. Martinelli, S. Nesseris, Constraints on the distance duality relation with standard sirens, Journal of Cosmology and Astroparticle Physics 2020 (12) (2020) 019.
[97] M. Martinelli, C. J. A. P. Martins, S. Nesseris, D. Sapone, I. Tutusaus, A. Avgoustidis, S. Camera, C. Carbone, S. Casas, S. Ilić, et al., Euclid: Forecast constraints on the cosmic distance duality relation with complementary external probes, Astronomy & Astrophysics 644 (2020) A80.
[98] M. Visser, Jerk, snap and the cosmological equation of state, Classical and Quantum Gravity 21 (11) (2004) 2603.
[99] M. Visser, Cosmography: Cosmology without the einstein equations, General Relativity and Gravitation 37 (2005) 1541–1548.
[100] S. Capozziello, V. Cardone, H. Farajollahi, A. Ravanpak, Cosmography in f (t) gravity, Physical Review D 84 (4) (2011) 043527.
[101] A. Bouali, H. Chaudhary, U. Debnath, A. Sardar, G. Mustafa, Data analysis of three parameter models of deceleration parameter in frw universe, arXiv preprint arXiv:2304.13137 (2023).
[102] A. Bouali, B. Shukla, H. Chaudhary, R. K. Tiwari, M. Samar, G. Mustafa, Cosmological tests of parametrization q= α-β h in f (q) flrw cosmology, International Journal of Geometric Methods in Modern Physics (2023) 2350152.
[103] A. Bouali, H. Chaudhary, A. Mehrotra, S. Pacif, Model-independent study for a quintessence model of dark energy: Analysis and observational constraints, arXiv preprint arXiv:2304.02652 (2023).
[104] A. Bouali, H. Chaudhary, S. Mumtaz, G. Mustafa, S. Maurya, Observational constraining study of new deceleration parameters in frw universe, Fortschritte der Physik (2023) 2300033.
[105] A. Bouali, H. Chaudhary, U. Debnath, T. Roy, G. Mustafa, Constraints on the parameterized deceleration parameter in frw universe, arXiv preprint arXiv:2301.12107 (2023).
[106] U. Alam, V. Sahni, T. Deep Saini, A. Starobinsky, Exploring the expanding universe and dark energy using the statefinder diagnostic, Monthly Notices of the Royal Astronomical Society 344 (4) (2003) 1057–1074.
[107] K. Bamba, S. Capozziello, S. Nojiri, S. D. Odintsov, Dark energy cosmology: the equivalent description via different theoretical models and cosmography tests, Astrophysics and Space Science 342 (2012) 155–228.
[108] V. Sahni, T. D. Saini, A. A. Starobinsky, U. Alam, Statefinder—a new geometrical diagnostic of dark energy, Journal of Experimental and Theoretical Physics Letters 77 (2003) 201–206.
[109] E. J. Copeland, M. Sami, S. Tsujikawa, Dynamics of dark energy, International Journal of Modern Physics D 15 (11) (2006) 1753–1935.
[110] M. Martinelli, C. J. A. P. Martins, S. Nesseris, D. Sapone, I. Tutusaus, A. Avgoustidis, S. Camera, C. Carbone, S. Casas, S. Ilić, et al., Euclid: Forecast constraints on the cosmic distance duality relation with complementary external probes, Astronomy & Astrophysics 644 (2020) A80.
[111] U. Alam, V. Sahni, T. Deep Saini, A. A. Starobinsky, Is there supernova evidence for dark energy metamorphosis?, Monthly Notices of the Royal Astronomical Society 354 (1) (2004) 275–291.
[112] V. Sahni, T. D. Saini, A. A. Starobinsky, U. Alam, Statefinder—a new geometrical diagnostic of dark energy, Journal of Experimental and Theoretical Physics Letters 77 (2003) 201–206.
[113] U. Alam, V. Sahni, T. Deep Saini, A. Starobinsky, Exploring the expanding universe and dark energy using the statefinder diagnostic, Monthly Notices of the Royal Astronomical Society 344 (4) (2003) 1057–1074.
[114] S. I. Vrieze, Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic)., Psychological methods 17 (2) (2012) 228.
[115] M. Tan, R. Biswas, The reliability of the akaike information criterion method in cosmological model selection, Monthly Notices of the Royal Astronomical Society 419 (4) (2012) 3292–3303.
[116] M. Rezaei, M. Malekjani, Comparison between different methods of model selection in cosmology, The European Physical Journal Plus 136 (2) (2021) 219.
[117] F. Arevalo, A. Cid, J. Moya, Aic and bic for cosmological interacting scenarios, The European Physical Journal C 77 (2017) 1–13.

[118] K. Tauscher, D. Rapetti, J. O. Burns, A new goodness-of-fit statistic and its application to 21-cm cosmology, Journal of Cosmology and Astroparticle Physics 2018 (12) (2018) 015.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205378

Credit Card Fraud Detection using Neural Embeddings and Radial Basis Network with a novel hybrid fruitfly-fireworks algorithm

Indu Singh[1], Nilesh Aditya[2], Pranjal Srivastava [3], Shivanshu Mittal[4], Taruvar Mittal[5], and Noel Vaibhav Surin[6]

[1,2,3,6]Department of Computer Science and Engineering, Delhi Technological University
[4]Department of Manufacturing Processes and Automation Engineering, Netaji Subhas University of Technology
[5]Department of Information Technology, Netaji Subhas University of Technology

Email: [1]indusingh@dtu.ac.in, [2]nileshaditya2000@gmail.com, [3]pranjaloct22@gmail.com,
[4]shivanshumittal31@gmail.com, [5]mittaltaruvar@gmail.com, [6]noelvaibhav@gmail.com

***Abstract*—Financial institutions are highly concerned about credit card fraud due to the growing frequency of deceitful activities linked with credit card transactions. To address this issue, a novel fraud detection system is proposed that combines advanced feature engineering and neural network architecture. The proposed approach. FFFW-RBN, employs a hybrid swarm intelligence model comprising the Fruitfly Optimization Algorithm and Fireworks Algorithm, (FFO-FWA) for feature selection, and the Neural Embeddings Generator in combination with a Shallow Radial Basis Network (RBF) for localized pattern classification. The proposed system achieves a 99.97% accuracy, 0.91 recall, and a precision of 0.94. The experimental results demonstrate that our methodology effectively and practically detects fraudulent credit card transactions. The paper presents a comprehensive evaluation of the system, including various performance metrics and a confusion matrix, highlighting its potential for real-world implementation.**

***Index Terms*—Credit Card Fraud, Metaheuristic optimisation, Neural Networks, Radial Basis Function Network, Imbalanced datasets**

I. Introduction

The rise of e-commerce and mobile internet technologies has resulted in a considerable advancement in online payment methods, such as credit cards, which have become increasingly popular [1]. In the present times, online transactions are fast, uncomplicated, and highly convenient. In case an individual uses a credit card that has been annulled, terminated, declared missing, or stolen, with the intention of deceiving to acquire any valuable commodity, it constitutes credit card fraud, which could potentially harm the cardholder's physical and psychological health.

Two kinds of Credit Card Fraud (CCF) can be recognized: Application fraud and Behavior fraud [2]. Credit Card Fraud (CCF) has worrisome effects on the entire world, costing businesses and people hundreds of millions of dollars yearly. According to a recent study, credit card fraud caused a loss of approximately $27.85 billion in 2018, which represents a 16.2% surge from the $23.97 billion lost in 2017. Furthermore, the report estimates that the losses will increase to $35 billion by 2023. [3]

Cybercriminals tend to prefer conducting their illicit activities online, where they can conceal their identity and location. The accuracy of CCF is significantly affected by the use of various testing, variable selection, and identification procedures. Data is constantly evolving, which means that actions that were previously considered legitimate may now be deemed fraudulent, and vice versa. The lack of reliable statistics is mainly attributed to financial institutions' concerns about their client's sensitive data and privacy violations, resulting in a scarcity of openly accessible data. The acceptance or denial of transactions happens within a fraction of a second, typically within milliseconds or microseconds of detecting the CCF. A problem arises when multiple transactions occur simultaneously and bear similarities to one another.

Therefore, prompt and effective identification of fraudulent conduct is vital. Given the variability of the issue depending on the context and industry, there is no single approach responsible for fraud detection that can be universally applied across all sectors. As a result, bespoke solutions are developed for each industry to aid in the reduction of fraud.

This paper introduces a novel framework for credit card fraud detection, incorporating a hybrid Fruitfly-Fireworks algorithm for feature selection, a neural embeddings generator for dense feature representation extraction, and a shallow Radial Basis Network for localized pattern classification. The proposed methodology is thoroughly evaluated using various performance metrics.

The key findings of this research are outlined below:

1) A novel Hybrid Fruitfly-Fireworks algorithm is proposed that incorporates enhanced exploration characteristics for feature selection.
2) A Neural Embeddings Generator based approach to extract dense feature representations for fraud detection is used for improved interpretability and dimensionality of the features.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:28 UTC from IEEE Xplore. Restrictions apply.

3) A Shallow Radial Basis Network is used for final classification that enhances the localized pattern discrimination in fraud detection.

Section II gives an overview of previous studies conducted on detecting credit card fraud, while Section III elaborates on the proposed methodology. Section IV presents the experimental outcomes and their analysis.

II. Literature Review

The detection of credit card fraud is a vital research field in the realm of information security. Customers and businesses may suffer financial losses as a result of deceitful actions in credit card transactions. In the early years, *Angus M Marshal et al. (2005)* [4] analyzed the motives and tactics of Internet-based identity theft and discussed how such incidents might be detected, investigated, and prevented in the future.

The study presented by *Jon T.S. Quah et al.(2008)* [5] focused on real-time fraud detection and proposed a novel approach to identify potential fraud cases by analyzing spending patterns. Self-organization maps were employed as a means of decoding, filtering, and analyzing customer behavior for fraud detection.

Rule-based systems are the earliest and most straightforward method for credit card fraud detection. *Suvasini Panigrahi et al. (2009)* [6] suggested a system that integrates three methods: Bayesian learning, Dempster-Shafer theory and rule-based filtering. The rule-based component filters incoming transactions based on predefined rules, while Dempster-Shafer's theory combines evidence from multiple sources to calculate the initial belief about each transaction. Bayesian learning, on the other hand, enables the system to learn from past transactions and update its beliefs accordingly. *Kang Fu et al.(2016)* [7] proposed a framework for fraud detection based on Convolutional Neural Networks (CNNs), which can identify complex patterns of fraudulent behavior through the analysis of labeled data.

Ishan Sohony et al. (2018) conducted a study that assessed the efficacy of a Random Forest and Neural Network fusion for the precise prediction of credit card fraud. Their research concluded that while Random Forest was better at detecting normal instances, Neural Network performed better at identifying fraudulent ones.

Vaishnavi Nath Dornadula et al. (2019) [8] explored an innovative approach, which involved the grouping of customers based on their transactional behavior and the extraction of behavioral patterns to develop a profile for each cardholder. The suggested method involved utilizing various classifiers to analyze the transactional data of three different groups, and generated rating scores for each classifier type.

In recent times, there has been a significant increase in the use of machine-learning techniques for detecting fraudulent credit card transactions. The objective of this literature review is to present an outline of these methods, with emphasis on ensembling and ML techniques. One such research proposed by *Dejan Varmedja et al. (2019)* [9] was aimed to evaluate and compare the efficiency of four widely-used ML algorithms in detecting fraudulent activities in credit card transactions. The assessed algorithms comprised of Random Forest, Logistic Regression, Multilayer Perceptron, and Naive Bayes. The research employed the SMOTE to tackle the problem of imbalanced data. The study discovered that among various algorithms tested, Random Forest algorithm produced the most precise outcomes in identifying fraudulent transactions.

Naoufal Rtayli et al. (2020) [10] conducted a comprehensive study and proposed a combined approach comprising three methods. One of these methods is the Support Vector Machine-Recursive Feature Elimination (SVM-RFE), which aims to extract the most effective predictive features and enhance the model's efficiency. The second method involves the SMOTE technique which is employed to convert a dataset that is imbalanced into one that is balanced. This is done in order to detect credit card frauds (CCFs) using a Random Forest Classifier (RFC). Additionally, the GridSearchCV method is used to optimize the parameters of the RFC, and it serves as a Hyper Parameter Optimization (HPO) solution.

Yvan Lucas et al. (2020) [11] introduced a novel approach for feature engineering using Hidden Markov Models (HMMs), which facilitated the integration of sequential knowledge in transactions by incorporating HMM-based features. As a result, a non-sequential classifier like Random Forest was able to efficiently incorporate sequential information during the classification process.

Taha et al. (2020) presented a fraud detection system for credit card transactions, which employs a Bayesian-based hyperparameter optimization and an optimized light gradient boosting machine (OLightGBM). The system includes dynamic parameter changes for adapting to new transactional behaviors and a feedback mechanism to address concept drift. The study evaluated various algorithms and found random forest, logistic regression, and decision tree to be the most effective for fraud detection using the proposed method.

In addition to traditional machine learning algorithms, researchers have also delved into deep learning methods for the identification of credit card frauds. *G. K. Arun et al. (2020)* [12] elaborated on the creation of a C-LSTM model for credit fraud detection, which encompasses data preprocessing and classification.

In another study by *Sumit Mishra et al. (2020)* [13], an autoencoder-based approach was proposed to convert transaction attributes into a feature vector of lower dimension, which was subsequently fed as input to a classifier in the second stage. *R. Asha et al.(2021)* [14] in their study, employed several machine learning algorithms, including Artificial Neural Network (ANN), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). To address the challenges associated with using an imbalanced dataset, they performed normalization, under-sampling, and data pre-processing.

Another such study was suggested by *Emmanuel Ileberi1 et al. (2022)* [15], where they used Genetic Algorithm (GA) to perform feature engineering in combination with different Machine Learning Algorithms.

Finally, researchers have also explored swarm intelligence

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:28 UTC from IEEE Xplore. Restrictions apply.

algorithms for credit card fraud detection. In a recent paper, *N. Prthenabhakaran et al. (2023)* [1] proposed the OCSODL-CCFD technique, which was used for the classification of fraudulent transactions involving credit cards. This technique consists of multiple subprocesses, including preprocessing, OCSO-based feature selection, the BiGRU classifier, and the CKHA-based hyperparameter optimizer

III. Proposed Methodology

This study proposes a novel approach for credit card fraud detection that consists of three key components: feature selection, neural embeddings generation, and classification. The feature selection step employs a hybrid fruitfly and fireworks algorithm to identify the most relevant features for the classification task. The selected features are then transformed into dense vector embeddings by a neural embeddings generator, which projects the feature space into a lower-dimensional, more informative space.

The resulting embeddings are fed into a shallow neural network with few dense layers, which captures the underlying patterns and relationships in the data. The final classification layer consists of a Radial Basis Network, which performs classification based on the spatial localities in the output of the neural network dense layers.

This approach offers several advantages over traditional methods, including improved interpretability of the features, more efficient computation due to the reduced dimensionality of the embeddings, and better classification performance due to the local nature of the Radial Basis Network.

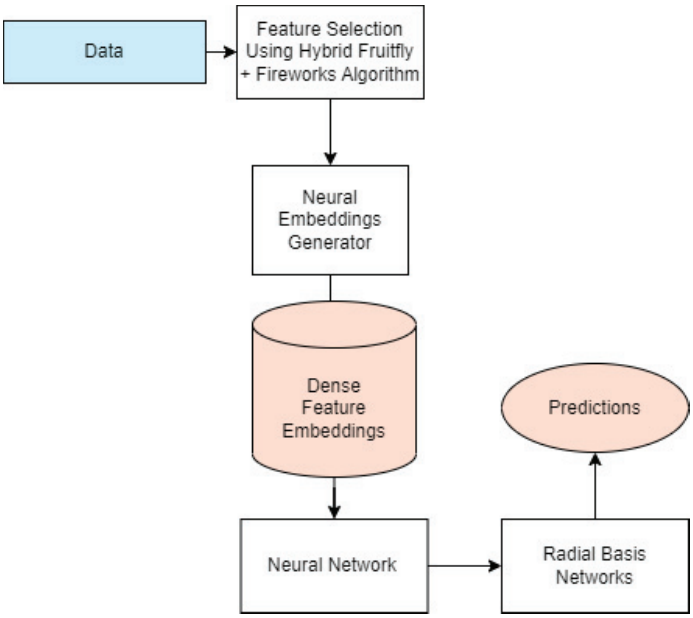


Fig. 1: Model Architecture

A. Hybrid Fruitfly-Fireworks Algorithm

Wen-Tsao Pan developed The fruit fly optimization algorithm [16], which is inspired by the foraging behavior of drosophila in their natural habitat. It is a population-based algorithm that mimics the foraging process of fruit flies, which rely on mutual information transfer via olfactory sense for migration of the entire population. The algorithm generates a population of fruit flies in the search space of a target problem and evolves the population to find the optimal solution.

The algorithm initializes the population with randomly generated individuals in the search space. Each individual has an x-axis and a y-axis representing its position. The algorithm assigns artificial search direction and search distance to the population, as shown in Equation 1 below, where L is the length of the search distance.

$$X_i = X_{axis} + \text{Rand}(L) \quad Y_i = Y_{axis} + \text{Rand}(L) \tag{1}$$

In the absence of knowledge regarding the location of food, the first iteration involves estimation of the distance from the origin (Dist), followed by the calculation of the smell concentration judgment value (S), which is the inverse of the distance. Subsequent iterations utilize the distance from the currently identified best candidate instead.

$$\text{Dist}_i = \sqrt{X_i^2 + Y_i^2} \tag{1}$$

$$V_i = \frac{1}{\text{Dist}_i} \tag{2}$$

The smell concentration judgment value (V_i) is utilized as an input to the Fitness function to determine the smell concentration ($Smell_i$) of each fruit fly's location.

$$\text{smell}_i = \text{Function}(V_i) \tag{3}$$

The population evolves by migrating in the direction of the best individual solution found so far. The algorithm records and preserves the individual optimal concentration and position coordinates using Equation 4 below, and the entire fruit fly population migrates in the optimal direction.

$$i^* = \underset{i}{\text{argmax}}(\text{smell}_i) \tag{4}$$

$$Smell^* = smell_{i^*} \quad X_{axis} = X_{i^*} \quad Y_{axis} = Y_{i^*} \tag{5}$$

However, a major disadvantage of FFAO is that the use of an elitist reproduction strategy in the fruit fly optimization algorithm may result in the swarm getting trapped in local extrema, thereby limiting its ability to effectively tackle complex, high-dimensional and nonlinear problems. Also, the smell concentration is not easily translatable to the problem of feature selection.

In our research, we address this issue by augmenting the exploration capabilities of the Fruitfly optimization algorithm. To achieve this, we have integrated the Fruitfly algorithm with the Fireworks algorithm [17]. This hybridization enables the Fruitfly algorithm to better explore the search space and overcome local optima. It also allows us to use the accuracy of the resulting model as the fitness function.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:28 UTC from IEEE Xplore. Restrictions apply.

Fireworks Algorithm (FWA) is a metaheuristic optimization algorithm inspired by fireworks explosions. In FWA, a firework represents a candidate solution and the explosion of a firework represents the generation of new candidate solutions. Each firework explodes to produce a set of sparks, and the quality of the firework determines the number of sparks it generates.

The method presented in this study employs sparks produced by fireworks to facilitate exploration of the search space and enhance the quality of candidate solutions in the fruitfly algorithm. The quantity of sparks generated is determined based on the performance of the firework and the best performance discovered thus far, as illustrated in Equation 6. This enables more thorough exploration of the search space near suboptimal candidates, thereby enhancing the quality of the search process.

$$sparks_i = N_{sparks} \cdot \frac{f_{\max} - f(x_i) + \phi}{\sum_{j=1}^{S} (f_{\max} - f(x_j)) + \phi} \tag{6}$$

Here, the parameter N_{sparks} controls the overall number of sparks generated via S fireworks. The optimum of the objective function achieved among the S fireworks is represented as $f_{\max}$, where $f(x_i)$ denotes the objective function value for $i = 1, 2, \ldots, S.$, and ϕ represents the constant used in the computer to avoid zero-division errors.

We also define the amplitude of perturbation L which controls the amplitude of spark generation. This is decayed with every update to the solution candidate as per the following Equation:

$$L = L.2^{-k} \tag{7}$$

where k is the number of updates to the solution.

We combine these to enhance the exploration by performing fireworks update on each individual fruitfly population member

B. Neural Embeddings Generator

Let's denote our input dataset as X, which has n samples and m features. We can represent the output of the neural embeddings generator as E, which has n rows and d columns, where d is the dimensionality of the embeddings.

The process of generating neural embeddings can be represented by a feedforward neural network with a hidden layer. The input layer of this network consists of m neurons, each corresponding to a feature in the input data X. The output layer contains d neurons, each corresponding to a dimension in the resulting embeddings.

Let's denote the matrix for neurons weight between input and the first layer as N_1, which has dimensions $d \times n$, where n is the number of neurons and d the dimensionality of the feature space. The bias vector for the hidden layer can be denoted as b_1, which has length n. Similarly, the matrix of neuron weights between the hidden layer and output is denoted as N_2, which has dimensions $n \times d$. The bias vector for the output layer can be denoted as b_2, which has length d.

Algorithm 1: Hybrid Fruit Fly - Fireworks Algorithm

Input: Population size N, maximum number of iterations T, problem dimension D, initial amplitude of perturbation L, Fitness function Φ, number of sparks S
Output: Best solution $\mathbf{x}^*$
Initialize population $\mathbf{X}$ randomly within the search space;
for $t = 1$ *to* T **do**
  for $i = 1$ *to* N **do**
    Generate artificial search direction $\mathbf{r}_i \sim \mathcal{U}(0,1)^D$;
    Calculate artificial step size $\alpha_i = L \cdot |\mathcal{N}(0,1)|$;
    Calculate new position $\mathbf{x}_i = \mathbf{x}_i + \alpha_i \cdot \mathbf{r}_i$;
    Calculate fitness $f_i = \Phi(\mathbf{x}_i)$;
  end
  Update X using fireworks exploration: $\mathbf{X} \leftarrow$ explore_fireworks$(\mathbf{X}, \Phi, L, S)$;
  Identify best solution $\mathbf{x}^*$ and its fitness f^*;
  for $i = 1$ *to* N **do**
    Calculate taste concentration value $\text{Smell}_i = f^* - f_i$;
  end
  Identify individual with lowest taste concentration $\text{Smell}_{\text{min}}$ and corresponding position $\mathbf{x}_{\text{min}}$;
  if $Smell_{min} > 0$ **then**
    Update best solution $\mathbf{x}^* = \mathbf{x}_{\text{min}}$;
    Update amplitude of perturbation $L = L/2$;
  end
end

Algorithm 2: Fireworks Exploration Update

Input : Candidate solutions X, fitness function Φ, amplitude of explosion L, number of sparks S
Output: Best solution after exploration x'
Intialize $X' = \ ;$
Calculate $f^* = \max(\Phi(x))$;
for $x_i \subset X$ **do**
  $s = S * \frac{y_{\max} - f(x_i) + \xi}{\sum_{j=1}^{n} (y_{\max} - f(x_j)) + \xi}$;
  for $j \leftarrow 1$ **to** s **do**
    Generate Gaussian step $g \sim \mathcal{N}(0,1)$;
    Calculate amplitude of spark $A = Lg$;
    Generate new spark $spark_j = x_i + A$;
  end
  Select the best spark as $spark^* = \text{argmax}_{spark_j}(\Phi(spark_j))$;
  $x_i \leftarrow spark^*$;
  $X' \leftarrow X' \bigcup x_i$;
end
return X'

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:28 UTC from IEEE Xplore. Restrictions apply.

The output E of the neural embeddings generator can be computed as follows:

$$E = \sigma(N_2 \cdot \sigma(N_1 \cdot X + b_1) + b_2) \tag{8}$$

where σ is an activation function, typically chosen to be a nonlinear function such as the sigmoid or ReLU. The activation function introduces nonlinearities into the model, allowing it to capture complex patterns in the data.

The weights and biases of the neural embeddings generator can be learned through backpropagation, using a loss function that measures the discrepancy between the predicted embeddings and the true embeddings. The specific loss function will depend on the task being performed, such as classification or regression.

Neural embeddings are useful in credit card fraud detection because they can project the high-dimensional feature space of transactions into a lower-dimensional space that captures the relevant patterns and relationships between features. This can reduce the dimensionality of the data while preserving important information, and capture non-linear relationships between features that are difficult to model with traditional techniques.

C. Neural Network

After generating the neural embeddings, The subsequent stage in the suggested approach is to feed them into a shallow neural network that reprojects them into a higher-dimensional space, which the RBF network can then compare. Specifically, we use a feedforward neural network with a few dense layers, followed by a final output layer with the same number of dimensions as the RBF network.

Let $f(\mathbf{x})$ denote the function computed by the neural network, where $\mathbf{x}$ represents the input embedding. The output of the i-th dense layer is computed as follows:

$$\mathbf{o}_i = \sigma(\mathbf{N}_i \mathbf{o}_{i-1} + \mathbf{b}_i) \tag{9}$$

where $\mathbf{N}_i$ is the matrix for neuron weights for layer i and $\mathbf{b}i$ is the bias vector for the i-th layer. $\mathbf{o}_{i-1}$ is the mapping obtained by previous layer (or the input embedding $\mathbf{x}$ for the first layer), and $\sigma(\cdot)$ denotes an activation function.

The final output layer maps the output of the last dense layer to a vector of the same dimensionality as the RBF network. This can be computed as follows:

$$\mathbf{y} = \mathbf{N}_{out} \mathbf{o}_L + \mathbf{b}_{out} \tag{10}$$

where $\mathbf{W}_{out}$ and $\mathbf{b}_{out}$ are the matrix for neuron weights and bias vector for the output layer, respectively, and $\mathbf{h}_L$ is the output of the last dense layer. The resulting output vector $\mathbf{y}$ can then be used as input to the RBF network for classification.

D. Radial Basis Networks

Radial basis Networks (RBNs) [18] are a type of artificial neural network (ANN) commonly used for classification and regression tasks. A Radial Basis Network consists of an input layer, a hidden layer with radial basis functions, and an output layer.

The hidden layer computes the output z_j of each of its J nodes using the radial basis function $\phi_j(\mathbf{x})$:

$$z_j = \phi_j(\mathbf{x}) = \exp\left(-\frac{|\mathbf{x} - \mu_j|^2}{2\sigma_j^2}\right), \tag{11}$$

where $\mathbf{x}$ is the input vector, μ_j is the mean for the jth radial basis function, and σ_j is the width modifier for the function.

The output layer computes the output $\hat{y}$ as a linear combination of the hidden layer outputs:

$$\hat{y} = \sum_{j=1}^{J} w_j z_j + b, \tag{12}$$

where w_j is the matrix of weights connecting the jth hidden node to the output node, and b is a bias term.

RBF networks exhibit a local nature, as the output of each hidden node is determined by a radial basis function that is only active in a local region of the input space, defined by its center and width parameters. This allows RBF networks to perform well on problems with spatial locality or clustered data, while also providing some regularization that can prevent overfitting and improve generalization.

IV. Experimentation and Results

A. Dataset

The credit card transaction dataset used in this study was obtained from the ULB Machine Learning Group [19], consisting of data from September 2013 and focusing on transactions made by cardholders from Europe. The dataset contains 284,807 transactions, of which only 492 are instances of fraud, resulting in a highly imbalanced dataset with the positive class representing a mere 0.172% of all transactions.

The dataset exclusively includes numerical input variables, which have undergone PCA transformation. The features V1, V2, ..., V28 correspond to the principal components obtained via transforming the input with PCA. This transformation was carried out to maintain user anonymity and confidentiality. The variable of interest is represented by the 'Class' feature, with a value of 1 indicating a transaction that is fraudulent, and 0 indicating otherwise.

B. Performance evaluation

For assessing our model, we perform a train test split of 70% and 30%, respectively as depicted in II.

TABLE II: Train-Test Split on Data

	Number of samples	Fraudulent	Non-Fraudulent
Train set	199364	344	199020
Test Set	85443	148	85295

In order to train the model, Weighted Binary cross-entropy loss was used, which measures the difference between predicted and actual class probabilities for each sample. It is defined as follows:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:28 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Comparison of our approach with other past approaches

Author	Technique Used	Performance Measure			
		Accuracy	Recall	Precision	F1 Score
Ishan Sohony et al. (2018) [20]	RF + NN	0.9995	0.8763	0.8585	0.8628
Varmedja et al. (2019) [9]	Random Forest + SMOTE	0.9996	0.8163	0.9638	0.8839
Misra et al. (2020) [13]	Autoencoder with MLP Classifier	0.9994	0.8015	0.8534	0.8266
Taha et al. (2020) [21]	Optimized Light Gradient Boosting Machine	0.9840	0.4059	0.9734	0.5729
Asha RB et al. (2021) [14]	ANN	0.9992	0.7619	0.8115	0.7859
	KNN	0.9982	0.0393	0.7142	0.0745
	SVM	0.9349	0.8976	0.9743	0.9343
Emmanuel Ileberi et al. (2022) [15]	GA-RF	0.9996	0.9912	0.8068	0.8895
Our Proposed Model (FFFW-RBN)	**Fruitfly-Fireworks + Radial Basis Networks**	**0.9998**	**0.94**	**0.96**	**0.9498**

The weighted binary cross-entropy loss function is defined as:

$$\delta = -\frac{1}{N}\sum_{i=1}^{N}[\omega_1 \cdot h_i \log(\hat{p_i}) + \omega_0 \cdot (1-h_i)\log(1-\hat{p_i})] \quad (13)$$

where N is the number of samples, h_i is the true binary label (either 0 or 1), and $\hat{p_i}$ is the predicted probability of the positive class. During training, we applied class weights of $\omega_1 = 10$ for the positive class (fraudulent transactions) and $\omega_0 = 1$ for the negative class (non-fraudulent transactions). To evaluate the performance of our trained model, we used several metrics commonly employed in binary classification tasks. Specifically, we used the metrics derived from confusion matrix, i.e., precision, recall, and F1 score, as well as the area under the precision-recall curve (AUPRC).

The confusion matrix is a table that summarizes the performance of a binary classification model. Based on these, we calculate the precision and recall for the model performance.

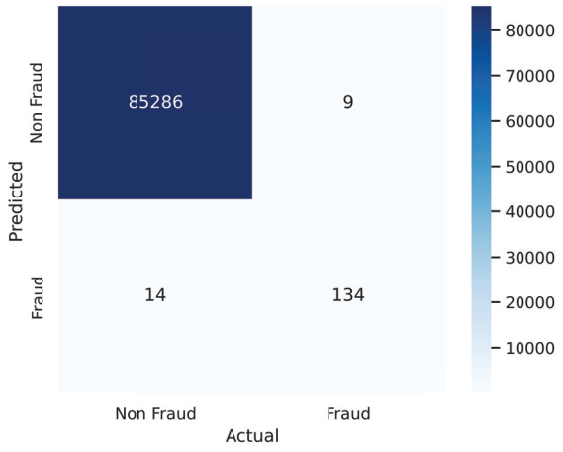


Fig. 2: Confusion Matrix

Precision is a performance metric that quantifies the proportion of correctly predicted positive cases out of all the cases that the model predicted as positive, and is defined as:

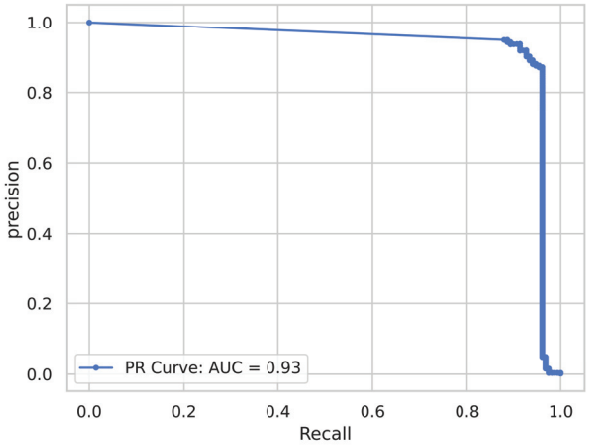


Fig. 3: Precision Recall Curve

$$Precision = \frac{True\ Pos.}{True\ Pos.\ +\ False\ Pos.} \quad (14)$$

Recall quantifies the proportion of true positive predictions in relation to all actual positive instances:

$$Recall = \frac{True\ Pos.}{True\ Pos.\ +\ False\ Neg.} \quad (15)$$

F1 score is a harmonic mean of precision and recall, and is useful for imbalanced datasets such as credit card fraud datasets.

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision\ +\ Recall} \quad (16)$$

In addition to the above metrics, we also evaluate the AUPRC, which is the area under the precision-recall curve. It is a measure of the trade-off between precision and recall. It is a useful metric when the dataset is highly imbalanced.

V. Conclusion and Future Work

In this paper, we proposed a novel fraud detection system for credit card transactions that utilizes hybrid swarm intelligence models and advanced feature engineering processes.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:28 UTC from IEEE Xplore. Restrictions apply.

Our results demonstrate that our methodology provides a practical and efficient means of detecting fraudulent transactions with high levels of accuracy, precision, and F1 score. The combination of the Fruitfly Optimization Algorithm and the Fireworks Algorithm for feature selection, and the Neural Embeddings Generator with a Shallow Radial Basis Function Network for pattern classification, shows promise for further development and use in real-world applications. In future work, the proposed methodology could be extended to other datasets and tested on larger credit card datasets with varying degrees of imbalance. The use of sequential models, such as recurrent neural networks or transformers, with the proposed feature engineering approach should also be explored as an area of research.

REFERENCES

[1] N. Prabhakaran, R. Nedunchelian, *et al.*, "Oppositional cat swarm optimization-based feature selection approach for credit card fraud detection," *Computational Intelligence and Neuroscience*, vol. 2023, 2023.

[2] X. Zhang, Y. Han, W. Xu, and Q. Wang, "Hoba: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture," *Information Sciences*, vol. 557, pp. 302–316, 2021.

[3] H. Tingfei, C. Guangquan, and H. Kuihua, "Using variational auto encoding in credit card fraud detection," *IEEE Access*, vol. 8, pp. 149841–149853, 2020.

[4] A. M. Marshall and B. C. Tompsett, "Identity theft in an online world," *Computer Law & Security Review*, vol. 21, no. 2, pp. 128–137, 2005.

[5] J. T. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," *Expert systems with applications*, vol. 35, no. 4, pp. 1721–1732, 2008.

[6] S. Panigrahi, A. Kundu, S. Sural, and A. K. Majumdar, "Credit card fraud detection: A fusion approach using dempster–shafer theory and bayesian learning," *Information Fusion*, vol. 10, no. 4, pp. 354–363, 2009.

[7] K. Fu, D. Cheng, Y. Tu, and L. Zhang, "Credit card fraud detection using convolutional neural networks," in *Neural Information Processing: 23rd International Conference, ICONIP 2016, Kyoto, Japan, October 16–21, 2016, Proceedings, Part III 23*, pp. 483–490, Springer, 2016.

[8] V. N. Dornadula and S. Geetha, "Credit card fraud detection using machine learning algorithms," *Procedia computer science*, vol. 165, pp. 631–641, 2019.

[9] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla, "Credit card fraud detection-machine learning methods," in *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)*, pp. 1–5, IEEE, 2019.

[10] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on svm-recursive feature elimination and hyper-parameters optimization," *Journal of Information Security and Applications*, vol. 55, p. 102596, 2020.

[11] Y. Lucas, P.-E. Portier, L. Laporte, L. He-Guelton, O. Caelen, M. Granitzer, and S. Calabretto, "Towards automated feature engineering for credit card fraud detection using multi-perspective hmms," *Future Generation Computer Systems*, vol. 102, pp. 393–402, 2020.

[12] G. Arun and K. Venkatachalapathy, "Convolutional long short term memory model for credit card detection," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1168–1172, IEEE, 2020.

[13] S. Misra, S. Thakur, M. Ghosh, and S. K. Saha, "An autoencoder based model for detecting fraudulent credit card transaction," *Procedia Computer Science*, vol. 167, pp. 254–262, 2020.

[14] R. Asha and S. K. KR, "Credit card fraud detection using artificial neural network," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 35–41, 2021.

[15] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the ga algorithm for feature selection," *Journal of Big Data*, vol. 9, no. 1, pp. 1–17, 2022.

[16] W.-T. Pan, "A new fruit fly optimization algorithm: taking the financial distress model as an example," *Knowledge-Based Systems*, vol. 26, pp. 69–74, 2012.

[17] Y. Tan and Y. Zhu, "Fireworks algorithm for optimization," in *Advances in Swarm Intelligence: First International Conference, ICSI 2010, Beijing, China, June 12-15, 2010, Proceedings, Part I 1*, pp. 355–364, Springer, 2010.

[18] M. D. Buhmann, "Radial basis functions," *Acta numerica*, vol. 9, pp. 1–38, 2000.

[19] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in *2015 IEEE symposium series on computational intelligence*, pp. 159–166, IEEE, 2015.

[20] I. Sohony, R. Pratap, and U. Nambiar, "Ensemble learning for credit card fraud detection," in *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pp. 289–294, 2018.

[21] A. A. Taha and S. J. Malebary, "An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine," *IEEE Access*, vol. 8, pp. 25579–25587, 2020.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:11:28 UTC from IEEE Xplore. Restrictions apply.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Cyberbullying Detection using Deep Learning Models in Bengali Language

Rohit Beniwal
Department of CSE
Delhi Technological University
Delhi, India
dr.rohitbeniwal@gmail.com

Shivam jha
Department of CSE
Delhi Technological University
Delhi, India
shivamjha1206@gmail.com

Smarth mehta
Department of CSE
Delhi Technological University
Delhi, India
mehtamaiden@gmail.com

Riyan dhiman
Department of CSE
Delhi Technological University
Delhi, India
riyandhiman14@gmail.com

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205775

***Abstract*—Social media has developed into a very powerful tool for teamwork, communication, and idea exchange. Due to the anonymity offered by these platforms, verbal abuse and cyberbullying cases have, sadly, proliferated globally. Researchers and academics from all around the world have conducted studies to create methods to automatically detect cyberbullying in response to this disturbing epidemic. Unfortunately, the research that is now accessible primarily focuses on more advanced languages, leaving low-resource languages like Bengali with a substantial research gap. As a result, in this study, we present a method for detecting cyberbullying in Bengali using Deep Learning Models such as Bidirectional Long Short-Term Memory (Bi-LSTM) and Convolutional Neural Network (CNN). Furthermore, we utilised the Bengali Cyber Bullying public dataset from kaggle.com to illustrate this. Also, when we examined the outcomes of the Bi-LSTM and CNN models, we discovered that CNN performed better. The precision, recall, and F1 score produced by the CNN model are on average 0.827, 0.850, and 0.855.**

Keywords—Bengali, Bi-LSTM, Cyber Bullying, Deep Learning, Machine Learning

I. INTRODUCTION

Bengali is an important language because it is the second most spoken language in India. Moreover, it is also the 7th most spoken language in the world [1]. It has a rich history along with its cultural heritage values. In recent years, Bengali has experienced a surge in popularity due to its widespread use on various social media sites. This has made it a useful language for people who are looking to connect with others from similar linguistic backgrounds. Bengali is the most commonly spoken language in several communities throughout many geographies, including Bangladesh, West Bengal in India, and in diaspora communities worldwide.

With the advent and growing popularity of social media sites such as Twitter, Instagram, Snapchat, Facebook, TikTok, Tumblr, etc., more and more people are able to express their opinions and thoughts freely. Due to their increasing popularity, particularly among teenagers, these social media platforms have become an essential part of our daily lives. Undoubtedly, the significant increase in the use of these platforms has certain drawbacks, and young teenagers are frequently exposed to a variety of behavioral and psychological risks.

One potential source of these hazards is cyberbullying in the form of powerful social attacks. Additionally, because social media platforms give users the choice to remain anonymous and conceal their identities, certain people may abuse their technical capabilities for their bad activities. Bullying that occurs more frequently over time makes things worse as well. The most dangerous behavior on social networking sites is now offensive language and abusive behavior intended to cause harm to others.

Cyberbullying may take many various forms, including offensive language, harassment, poisonous remarks, hate speech, sexism, racism, and other types of discrimination. Anxiety, despair, self-harm, social and emotional confusion, and even suicidal thoughts or attempts are all frequent negative effects of receiving abusive texts. The Pew Research Center found that teenage girls experienced the worst forms of cyberbullying and that around 59% of US users of social media platforms had encountered it [2].

Thus, the epidemic of cyberbullying has spread to every country. As a result, several international preventive and intervention strategies have been developed to enhance online users' safety and reduce such heinous cyberbullying. Hostile text detection is seen as a challenging and complex process even for humans, aside from the victims, because of the variations in the language used. Similarly, cyberbullying detection in Bengali has several challenges. Among those challenges, the first is the lack of sufficient data. One of the primary challenges is the lack of sufficient data to train machine learning models for cyberbullying detection. The Bengali language has a comparatively small amount of training data compared to other languages, making it difficult to create models that can accurately detect cyberbullying. The second is language complexity. Bengali is a complex language with multiple variations in grammar, syntax, and vocabulary. This complexity makes it challenging to identify the context and tone of messages, which is essential for detecting cyberbullying accurately. Slang and acronyms come in third place. Slang and acronyms are third on the list. When cyberbullying occurs, uncommon Bengali slang and acronyms are usually employed. These non-standard sentences may be difficult to recognize and analyze, which might lead to inaccurate detection. And fourth is cultural context. When creating models for the identification of cyberbullying, the cultural background of the Bengali language must be taken into consideration. What may be considered offensive or bullying in one culture may not be

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:02 UTC from IEEE Xplore. Restrictions apply.

the same in another, making it challenging to create a universal detection system. Lastly, the fifth is code-switching. Many Bengali speakers frequently switch between Bengali and English, particularly in digital communications. This code-switching can make it difficult to detect cyberbullying since models must be trained to recognize both languages.

In order to overcome the challenges of cyberbullying detection in Bengali, we provide "Deep Learning models such as Bi-LSTM and CN" [3] in this paper. We then evaluate the results using different metrics like the f1-score, recall, and precision.

The remaining paper is outlined as follows: Section 2 discusses related research on cyberbullying detection in Bengali. Section 3 describes the proposed methodology, whereas Section 4 summarizes the results and analysis. Lastly, Section 5 concludes the research paper by offering future research suggestions.

II. Related work

In today's society, there is increasing concern about cyberbullying, particularly among young people who use the internet. This term refers to a pattern of behavior based on electronic communication used to threaten, harass, or intimidate another person. Researchers are actively investigating various strategies to identify and stop cyberbullying due to the issue receiving more attention in recent years. Teenagers who use the internet today are particularly vulnerable to cyberbullying, also known as hate speech, cyber aggression, and toxic speech. Victims may suffer from serious consequences such as "low self-esteem, anxiety, depression, hopelessness and in some cases, a lack of motivation to be alive, ultimately resulting in the death of a victim" [4].

"In the field of machine learning and deep learning" [5], several studies have been conducted to classify comments as cyberbullying or not. Taawab et al. [6] employed various machine learning algorithms, such as decision trees, logistic regression, and recurrent neural networks, to classify abusive transliterated Bengali comments. To build their model, they created a custom dataset consisting of only 1,300 transliterated Bengali words, as there was no publicly available sized transliterated Bengali dataset available. They found that logistic regression with a count vectorizer achieved the highest accuracy and F1 score, at 85.76% and 85.70%, respectively.

In a study on deep learning methods, G and Uma [7] found that the Gated Recurrent Unit (GRU) had the highest accuracy 95.47% for identifying whether or not comments indicated cyberbullying. In another study, Emon et al. [8] tested several algorithms, such as Logistic Regression, Multinomial Naive Bayes, and "Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM)" [9] cells for classifying abusive Bengali text. They gathered data from the comment sections of various social media websites and internet resources. The dataset included both positive and neutral data as two separate classifications. They found that the Recurrent Neural Network (RNN) had the highest "precision, recall, and F1 score" [10], achieving 0.83, 0.82, and 0.82, respectively.

Altaf et al. [11] conducted a study on "cross-domain sentiment analysis on the Urdu language using" [11] deep learning and machine learning classifiers. They achieved an accuracy of 77%, a precision of 83%, a recall of 68%, and an F1 score of 74%.

In conclusion, the literature provides evidence that machine learning and deep learning can effectively detect and prevent cyberbullying in regional languages. Further research is needed to enhance the accuracy of the above-discussed algorithms and make them more robust in different languages and domains. Therefore, in this research paper, we provide cyberbullying detection using "Deep Learning models such as Bi-LSTM and CNN" [3] in the Bengali language, where we further enhance the F1 score as "compared to the state-of-the-art" [13].

III. METHODOLOGY

The methodology for cyberbullying detection using Deep Learning models in Bengali is built by evaluating it on a public dataset with the help of different standard performance metrics. The following Fig. 1 shows the research methodology.

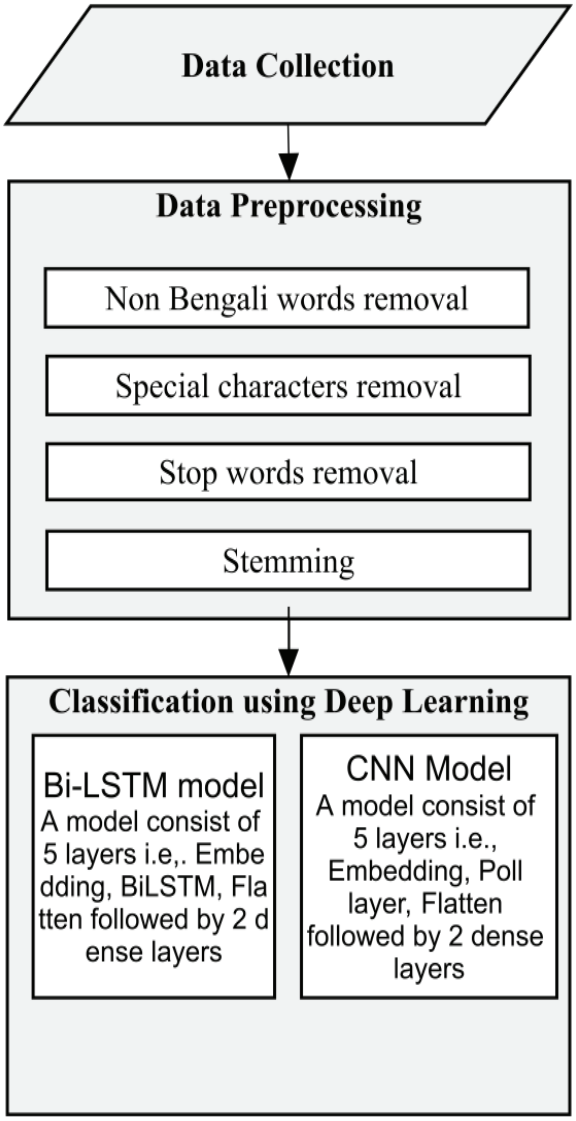


Fig. 1. The research methodology

The subsequent subsections explain the proposed research methodology used for cyberbullying detection using Deep Learning in the Bengali language.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:02 UTC from IEEE Xplore. Restrictions apply.

A. Dataset

The following Fig. 2 shows the distribution of the dataset.

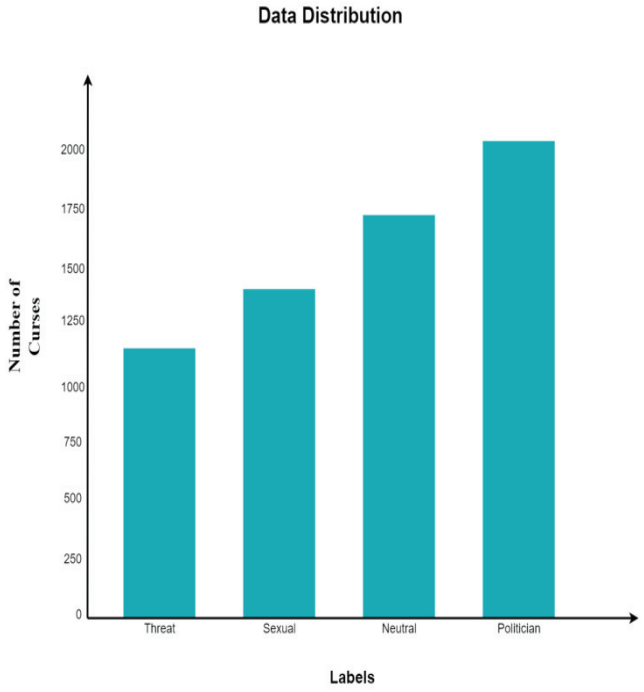


Fig. 2. Dataset distribution

To demonstrate the efficiency of the proposed research approach, we considered the Bengali Cyber Bullying public dataset available at kaggle.com [14]. This dataset was developed by Faisal using web scraping on different social media platforms, such as YouTube, Facebook, and Twitter. The dataset has more than 6000 comments, which are distributed among political, sexual, threat, and neutral domains. The dataset has both labeled instances of cyberbullying-positive as well as cyberbullying-negative cases.

B. Data preprocessing

Data preprocessing is carried out to eliminate noise from the dataset. Here, data preprocessing includes four steps. In the first step, we removed the non-Bengali words from the dataset. In the second step, we removed special characters such as white spaces, digits, and others like {, *, &, etc., which had no specific meaning [15]. In the third step, we removed the stop words. "Stop words are common English words such as the, am, their, which do not influence the semantics of the review" [15]. Finally, in the fourth step, we performed stemming. Stemming in Natural Language Processing (NLP) is the process of reducing words to their root or base form, also known as a stem, by removing prefixes or suffixes.

C. Classification using Deep Learning Models

This sub-section discusses the proposed Deep Learning models for cyberbullying detection in Bengali. Therefore, for cyberbullying detection, we use "Deep Learning models such as Bi-LSTM and CNN" [3] models.

1) Classification using Bi-LSTM

The following Fig. 3 shows the proposed Bi-LSTM model for classifying cyberbullying cases.

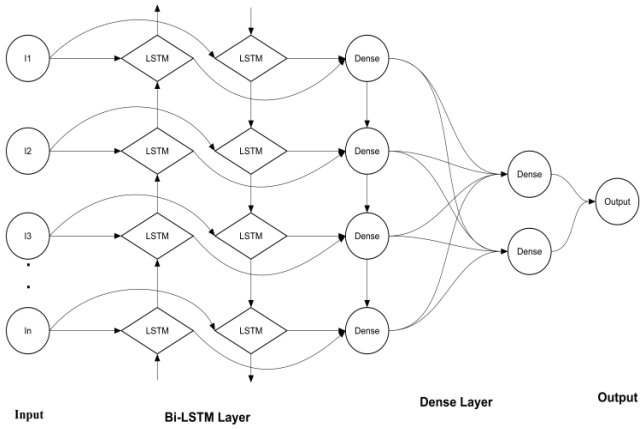


Fig. 3. Bi-LSTM Model

The Bi-LSTM model consisted of a total of five layers. The first layer was responsible for embedding and involved taking the integer-encoded vocabulary and searching for the embedding vector for each word-index. The output from this layer was then passed to the second layer, which was the Bi-LSTM layer. The Bi-LSTM layer learned the sequence information both in the forward and backward directions. Subsequently, the output from the Bi-LSTM layer was passed to the dropout layer, which eliminated the impact of certain neurons. Finally, the last two layers were dense layers, with the first layer using 'Relu' as the activation function, and the second layer using 'Softmax' as the activation function for the final output layer.

2) Classification using Convolutional Neural Network

The following Fig. 4 shows the proposed Convolutional Neural Network model for classifying cyberbullying cases.

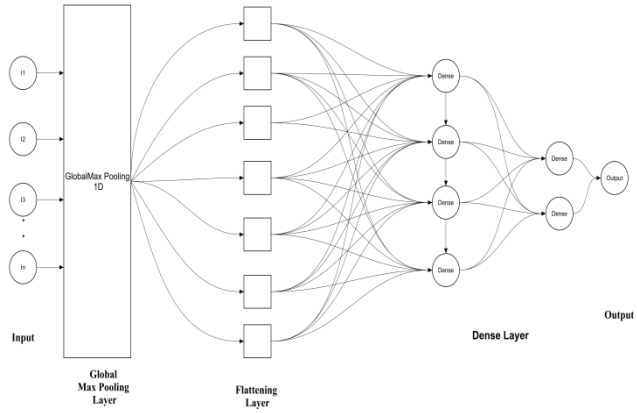


Fig. 4. The CNN Model

The CNN model consisted of a total of five layers. The only difference between the CNN and Bi-LSTM models was in the second and third layers. In the CNN model, the second layer was a global max pooling 1D layer, which downsampled information by taking the maximum over the time dimension. The output from this layer went to a flattening layer, which was used "to convert all the resultant 2-D arrays from pooled feature maps into a single linear vector" [7].

IV. RESULT AND ANALYSIS

The proposed methodology in this study is implemented using the Python programming language. In addition, the following performance indicators were used to evaluate the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:02 UTC from IEEE Xplore. Restrictions apply.

classification results: "precision, recall, and F1 score" [10]. "Precision measures how many of the samples predicted as positive are actually positive. Thus, it is defined as the ratio of true positive (TP) values to the sum of all predicted positive values (TP and FP).

Precision =True Positive / (True Positive + False Positive)

Recall measures how many of the positive samples are captured by the positive predictions. Therefore, it is defined as the ratio of true positive (TP) values to the summation of all actual positive values (TP and FN).

Recall = True Positive / (True Positive + False Negative)

F1-Score is defined as the harmonic mean of precision and recall" [16].

F1 score = 2 × Precision × Recall / (Precision + Recall)

A. *Implementation details for data preprocessing phase*

Table I presents an overview of the Python libraries and functions that are utilized for the preprocessing of data.

TABLE I. LIBRARIES AND FUNCTIONS USED FOR DATE PREPROCESSING

S.NO.	Preprocessing Task	Python Library/ Function Used
1.	Non-Bengali words removal	Python lower() function
2.	Special characters removal	BNLP Library
3.	Stop words removal	BNLP Library
4.	Stemming	Bangla Stemmer Library

B. *Implementation details and classification results using Deep Learning Models*

We utilized the Python programming language, along with the TensorFlow and Keras libraries, to develop our proposed Deep Learning models, including the Bi-LSTM and CNN models, for cyberbullying detection in Bengali. The following two sub-sections expound the classification results of the Bi-LSTM and CNN models.

1) *Implementation details and classification results of the Bi-LSTM model*

The proposed Bi-LSTM model was implemented using the Python programming language. The following Table II describes the various settings used for the model.

TABLE II. DESCRIPTION OF EACH LAYER IN THE BI-LSTM MODEL

Layer	Output Shape
Embedding	(None, None, 64)
Bidirectional – LSTM	(None, 128)
Dropout	″(None, 128)
Dense (activation='relu')	(None, 64)
Dense(activation='softmax')	(None″, 4)

The following Table III shows the classification results of the Bi-LSTM model.

TABLE III. PRECISION, RECALL, F1-SCORE FOR BI-LSTM MODEL

Class name	Precision	Recall	F1-Score
Neutral	0.734	0.817	0.773
Political	0.809	0.886	0.845
Threat	0.818	0.823	0.820
Sexual	0.889	0.652	0.752

The average "precision, recall, and F1 score" [10] for Bi-LSTM model are 0.812, 0.794, and 0.801.

2) *Implementation details and classification results of the CNN model*

The proposed CNN model was implemented using Python programming language. The following Table IV describes the various setting used for the model.

TABLE IV. DESCRIPTION OF EACH LAYER IN CNN MODEL

Layer	Output Shape
Embedding	(None, 669, 128)
CNN (Global max pooling 1D)	(None, 128)
Flatten	″(None, 128)
Dense(activation='relu')	(None, 64)
Dense(activation='softmax')	(None″, 4)

The following Table V shows the classification results of the CNN model.

TABLE V. PRECISION, RECALL, F1-SCORE FOR CNN MODEL

Class Name	Precision	Recall	F1-Score
Neutral	0.806	0.839	0.822
Political	0.874	0.905	0.889
Threat	0.876	0.833	0.854
Sexual	0.873	0.824	0.845

The average "precision, recall, and F1 score" [10] for Bi-LSTM model are 0.827, 0.850, and 0.855.

C. *Comparative analysis between classification results of Bi-LSTM vs CNN models*

The following Fig. 5 shows the comparative analysis between the classification results of the Bi-LSTM and CNN models.

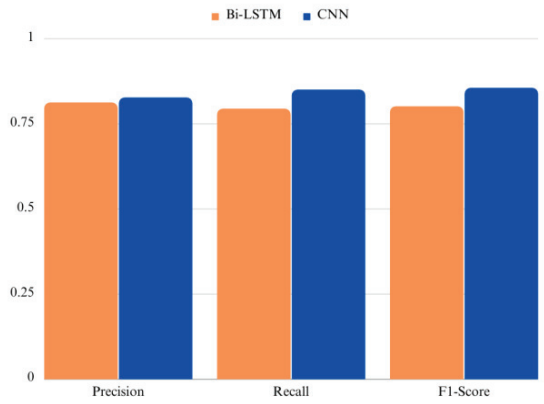


Fig. 5. Comparison of the Bi-LSTM and CNN model.

In Fig. 5, the orange bar displays the Bi-LSTM model's classification outcomes, whereas the blue bar displays the CNN model's outcomes. This figure makes it very evident that the CNN model's performance outperformed that of the Bi-LSTM model. The Bi-LSTM model has an average precision of 0.812, a recall of 0.794, and an F1 score of 0.801. On the other hand, the CNN model has an average precision of 0.827, a recall of 0.850, and an F1 score of 0.855.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:02 UTC from IEEE Xplore. Restrictions apply.

V. CONCLUSION AND FUTURE SCOPE

In this research paper, we provided a cyberbullying detection approach using "Deep Learning Models such as Bi-LSTM and CNN" [3] in the Bengali Language. We utilized the Bengali Cyber Bullying public dataset from kaggle.com to illustrate this. Also, when we examined the outcomes of the Bi-LSTM and CNN models, we discovered that CNN performed better. The "precision, recall, and F1 score" [10] for the CNN model were 0.827, 0.850, and 0.855, respectively.

There are several directions we might go in order to improve the model's accuracy in subsequent studies. One such method is ensemble modeling [17], which makes predictions by combining numerous models, greatly improving accuracy. Including strategies like word2vec or rapid tag-like vector embedding methods can also help the model perform better. Incorporating more data and fine-tuning the models for the Bengali language could also lead to improved performance. Furthermore, adding contextual information, such as user profiles and interactions, could enhance the accuracy of the models in identifying instances of cyberbullying. Additionally, exploring other techniques, such as transformer-based models, which have shown to be successful in various natural language processing tasks, could also lead to improvements.

In conclusion, our research has "demonstrated the effectiveness of deep learning models in" [18] identifying cyberbullying in the Bengali language and has highlighted the need for further research in this under-researched area. The developed models have the potential to be used in practical applications to detect and mitigate cyberbullying in the Bengali language.

REFERENCES

[1] "The most spoken languages worldwide in 2022." https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/ (accessed Apr. 07, 2023).

[2] M. Anderson, "A Majority of Teens Have Experienced Some Form of Cyberbullying." https://www.pewresearch.org/internet/2018/09/27/a-majority-of-teens-have-experienced-some-form-of-cyberbullying/ (accessed Apr. 07, 2023).

[3] Z. Chen, X. Ju, H. Wang, and X. Chen, "Hybrid Multiple Deep Learning Models to Boost Blocking Bug Prediction," Journal of Internet Technology, vol. 23, no. 5, pp. 1099–1107, 2022, doi: 10.53106/160792642022092305018.

[4] A. Dewani, A. Memon, and S. Bhatti, "Cyberbullying detection: advanced preprocessing techniques & deep learning architecture for Roman Urdu data", doi: 10.1186/s40537-021-00550-7.

[5] A. Tiwari, R. Dave, and M. Vanamala, "Leveraging Deep Learning Approaches for Deepfake Detection: A Review." Apr. 2023.

[6] A. Al Taawab, L. Tasnia, M. Dhar, and M. H. K. Mehedi, "Transliterated Bengali Comment Classification from Social Media," in 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), IEEE, Sep. 2022, pp. 365–371. doi: 10.1109/R10-HTC54060.2022.9929514.

[7] K. G. Apoorva and D. Uma, "Detection of Cyberbullying Using Machine Learning and Deep Learning Algorithms," 2022 2nd Asian Conference on Innovation in Technology, ASIANCON 2022, 2022, doi: 10.1109/ASIANCON55314.2022.9908898.

[8] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das, and T. Mittra, "A Deep Learning Approach to Detect Abusive Bengali Text," 2019 7th International Conference on Smart Computing and Communications, ICSCC 2019, Jun. 2019, doi: 10.1109/ICSCC.2019.8843606.

[9] P. L. Brockett, L. L. Golden, J. Jang, and C. Yang, "A Comparison of Neural Network, Statistical Methods, and Variable Choice for Life Insurers' Financial Distress Prediction," Journal of Risk and Insurance, vol. 73, no. 3, pp. 397–419, Sep. 2006, doi: 10.1111/J.1539-6975.2006.00181.X.

[10] M. P. S. Bhatia, A. Kumar, and R. Beniwal, "An Optimized Classification of Apps Reviews for Improving Requirement Engineering," Recent Advances in Computer Science and Communications, vol. 14, no. 5, pp. 1390–1399, Jul. 2019, doi: 10.2174/2213275912666190716114919.

[11] A. Altaf et al., "Deep Learning Based Cross Domain Sentiment Classification for Urdu Language," IEEE Access, vol. 10, pp. 102135–102147, 2022, doi: 10.1109/ACCESS.2022.3208164.

[12] M. Agüero-Torales, J. Abreu, and A. López-Herrera, "Deep learning and multilingual sentiment analysis on social media data: An overview," Appl Soft Comput, p. 107373, Apr. 2021, doi: 10.1016/j.asoc.2021.107373.

[13] O. Eljundi et al., "hULMonA: The Universal Language Model in Arabic," Proceedings of the Fourth Arabic Natural Language Processing Workshop, pp. 68–77, 2019, doi: 10.18653/V1/W19-4608.

[14] M. Faisal, "Bengali Cyber Bullying Dataset." https://www.kaggle.com/datasets/moshiurrahmanfaisal/bangla-cyber-bullying-dataset (accessed Apr. 07, 2023).

[15] W. Maalej and H. Nabil, "Bug report, feature request, or simply praise? On automatically classifying app reviews," in 2015 IEEE 23rd International Requirements Engineering Conference, RE 2015 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Nov. 2015, pp. 116–125. doi: 10.1109/RE.2015.7320414

[16] Kumar Akshi, Beniwal Rohit, and Jain Dipika, "Personality Detection using Kernel-based Ensemble Model for leveraging Social Psychology in Online Networks," ACM Transactions on Asian and Low-Resource Language Information Processing, May 2022, doi: 10.1145/3571584.

[17] K. S. Alam, S. Bhowmik, and P. R. K. Prosun, "Cyberbullying detection: An ensemble based machine learning approach," Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021, pp. 710–715, Feb. 2021, doi: 10.1109/ICICV50876.2021.9388499.

[18] Kumar, S., Kumar, A., Mallik, A., & Singh, R. R. (2023). OptNet-Fake: Fake News Detection in Socio-Cyber Platforms Using Grasshopper Optimization and Deep Neural Network. *IEEE Transactions on Computational Social Systems*, 1–10. https://doi.org/10.1109/TCSS.2023.3246479

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:13:02 UTC from IEEE Xplore. Restrictions apply.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-16286-w

Deep learning algorithms for person re-identification: sate-of-the-art and research challenges

Ankit Yadav[1] · **Dinesh Kumar Vishwakarma**[1]

Received: 8 September 2021 / Revised: 5 July 2023 / Accepted: 6 July 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Person re-identification has received a lot of attention from the research community in recent times. Due to its vital role in security based applications, person re-identification lies at the heart of research relevant to tracking robberies, preventing terrorist attacks and other security critical events. While the last decade has seen tremendous growth in re-id approaches, very little review literature exists to comprehend and summarize this progress. This review deals with the latest state-of-the-art deep learning based approaches for person re-identification. While the few existing re-id review works have analysed re-id techniques from a singular aspect, this review evaluates numerous re-id techniques from multiple deep learning aspects such as deep architecture types, common Re-Id challenges (variation in pose, lightning, view, scale, partial or complete occlusion, background clutter), multi-modal Re-Id, cross-domain Re-Id challenges, metric learning approaches and video Re-Id contributions. This review also includes several re-id benchmarks collected over the years, describing their characteristics, specifications and top re-id results obtained on them. The inclusion of the latest deep re-id works makes this a significant contribution to the re-id literature. Lastly, the conclusion and future directions are included.

Keywords Person Re-Identification · Deep Learning · Convolutional Neural Network · Feature Extraction & Fusion

1 Introduction

Security and surveillance based applications in computer vision have gained tremendous popularity in recent times. Currently, security surveillances record videos and images whose analysis requires manual human interaction. Looking into yesterday's robbery can be challenging as it involves a manual search through twenty hours of surveillance videos by humans prone

✉ Dinesh Kumar Vishwakarma
dvishwakarma@gmail.com

Ankit Yadav
ankit4607@gmail.com

[1] Biometric Research Laboratory, Department of Information Technology, Delhi Technological University, Bawana Road, Delhi 110042, India

to fatigue and making errors. The problem quickly becomes infeasible as the time span of recorded media under analysis is increased. The development of machine learning and later, deep learning approaches have opened a wide range of advanced possibilities that could lead to safer homes, offices, neighbourhood, bus stops, airports etc. The idea that machines can be taught to identify individuals of interest is a promising step towards a more secure environment.

Person re-identification means to find a person of interest in a query image/video from a large collection of recorded images and/or videos. While machine learning algorithms played crucial part in the early days, re-id approaches have made significant improvements with the rise of deep learning based systems [170]. Several deep learning based approaches proposed in recent years have boosted the matching accuracy results [149], significantly outperforming the handcrafted feature based machine learning algorithms [110, 126, 143]. Since deep learning models require a large number of training samples, recent years have also witnessed the collection of several medium to large-scale re-id datasets for training and testing different deep based re-id approaches.

The efficiency and real-time performance of person reidentification tasks were severely hindered by the fact that previous techniques relied on the manual extraction of distinctive features. Deep learning architectures have been used in nearly all current research in this field. There has been a surge in the number of published studies on the topic of human reidentification over the years. Some researchers have proposed various loss functions too to enhance the models' ability to extract the discriminative features. Additionally, others have introduced various attention modules to focus on important details of a person's body for the purpose to acquire more robust features of pedestrians. By merging both the local and global characteristics of a person, multiple research contributions improved the final representation of features.

In contrast, generative adversarial training was frequently employed for the tasks of person reidentification because of GAN's outstanding results in both the generation of images and feature acquisition. Some studies have found that video clips can effectively fuse more information than single-frame images due to the mutually beneficial temporal and spatial signals present in both. To simulate relationships in pedestrian pictures as well as to acquire more selective and reliable attributes, graph convolutional network-based approaches have recently emerged.

1.1 Motivation

Building a system to reidentify people is driven by a variety of different interests and priorities. The primary motivation to build person re-identification systems is as follows:

- *Safety:* There is an absolute requirement for the installation of person identification systems in all public areas, including shopping malls and airports, to maintain safety and put a stop to any criminal activity that may occur.
- *Scrutiny:* Monitoring people in public areas with the use of a person reidentification system can assist law enforcement organizations in the identification of potentially suspicious behaviors.
- *Marketing:* Using a person reidentification system, retailers can monitor customer activity to learn more about customer preferences and make more informed marketing decisions.
- *Traffic Regulation:* For traffic control, a person reidentification system may be utilized to enable tracking of passengers as well as automobiles.
- *Personalized Consumer Experience:* Identifying a given person allows for a custom and personalized shopping experience online since offers and recommendations can be arranged around individual consumer interests.

In general, these systems can serve businesses, government agencies, and the general public in a wide range of different ways, which is why they are becoming increasingly popular.

2 Research methodology

This section details the major contributions of this review, techniques followed in preparing this review and a comparison with some of the existing deep Re-Id reviews/surveys.

2.1 Contributions of this review

This review studies deep learning based person re-identification. While person re-identification is not a new topic in the research community, deep learning based methods have become increasingly popular due their tremendous success in various computer vision domains. Hence, it is natural that deep learning based methods take lead in the Re-Id problem due to their superior feature learning capability when compared to hand-crafted feature based methods. The focus of this review is to conduct an exhaustive study of the recent deep learning based Re-Id methods, specifying the various benchmark datasets and the classification of various image and video based deep methods.

Figure 1 presents the organization of this review. Specifically, Section 1 presents the Introduction to person re-identification. Then, Section 2 highlights the

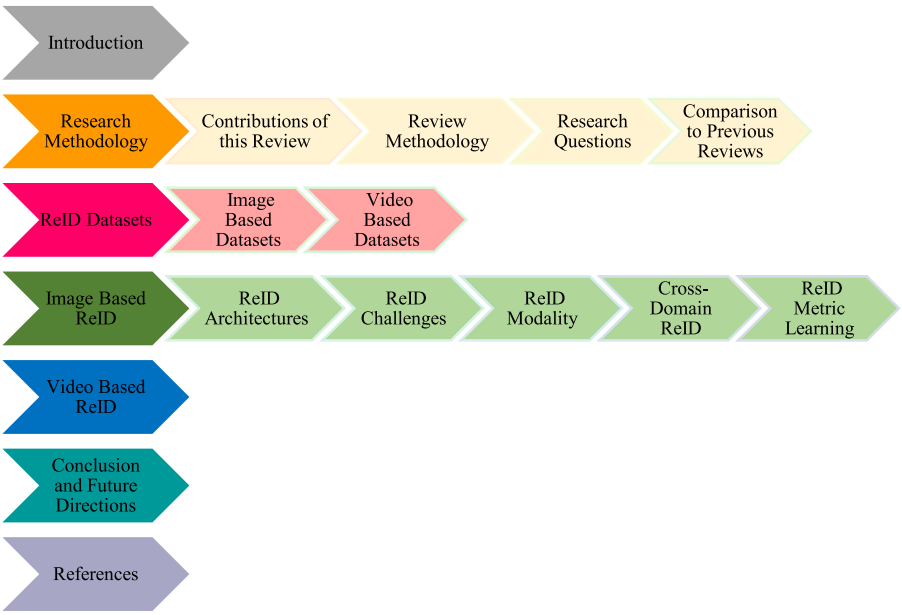


Fig. 1 Organization of this review

importance of this review by mentioning its major contributions, the review methodology followed, research questions being addressed by this review and finally a comparison with existing person re-identification reviews. Section 3 talks about the major publicly available benchmark datasets for person re-identification. Section 4 explores the different kinds of image-based person reid methods. Section 5 discusses video-based person re-id research papers. Finally, the conclusion and future work are discussed in Section 6.

Recent years have witnessed a sharp increase in the number of deep learning based Re-Id approaches and this review cites the latest developments in deep learning based Re-Id research. Figure 2 describes the taxonomy of deep learning based Re-Id methods included in this review. Re-Id methods have been categorized into image-based and video-based methods. Image-based contributions have been explored from numerous aspects such as architectures involved, common visual challenges, modality specific methods, cross-domain approaches and metric learning methods.

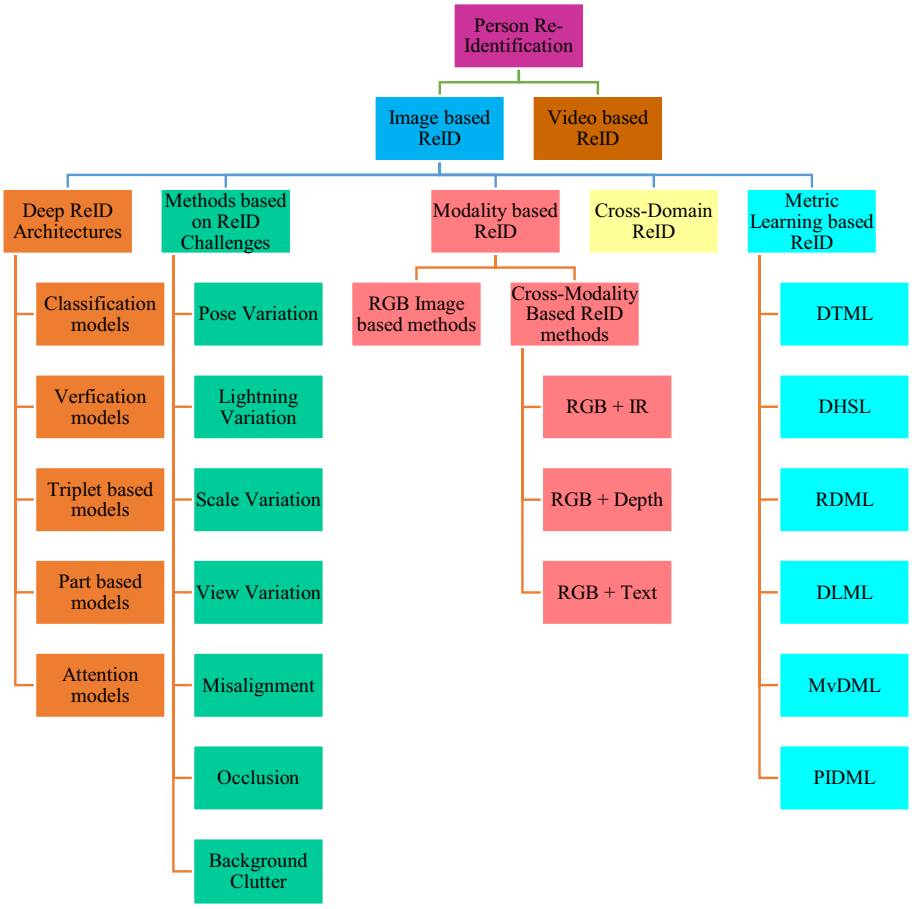


Fig. 2 Taxonomy of Re-Id methods

The major contributions of this review are as follows:

- Provides a comprehensive review of deep learning based person re-identification methods.
- Conducts an exhaustive study of deep Re-Id methods (Tables 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11) by describing the "key ideas" behind numerous approaches mentioned.
- Since deep learning methods have gained popularity in recent years, this exhaustive review automatically incorporates the most recent contributions to deep learning based Re-Id approaches.
- Details several image based and video based benchmark datasets, specifying their technical specifications, challenges posed by samples and top results reported on them.
- Analyses deep learning based methods from several crucial aspects like architecture types, loss functions, Re-Id challenges, data modalities, cross-domain approaches and metric learning methods, helping the readers to understand and appreciate deep Re-Id from multiple perspectives.
- Explores the growing popularity of video based deep Re-Id methods that combine temporal data providing important motion cues with the usual visual characteristics.
- This review helps readers to gain a comprehensive and exhaustive understanding of the recent deep Re-Id contributions by categorically analysing contributions from numerous perspectives like architecture, challenges, modality, cross-domain approaches, metric learning and video based methods.

Table 1 Comparison of this current review with the existing deep learning based Re-Id reviews/surveys. rows in green indicate theoretical analysis while orange rows demonstrate tabular information

Deep Re-Id Reviews/Surveys		Wang et al. [6]	Almasawa et al. [7]	Wu et al. [8]	Islam [9]	Current Review
Year of Publication		2018	2019	2019	2020	-
Year of Latest Citation		2018	2019	2018	2020	2020
Re-Id Datasets	Theory		✓	✓	✓	✓
	Table		✓		✓	✓
Image & Video Re-Id Methods	Theory		✓			✓
	Table		✓			✓

Deep Re-Id Reviews/Surveys			Wang et al. [6]	Almasawa et al. [7]	Wu et al. [8]	Islam [9]	Current Review
Deep Re-Id Analysis	Architecture Types	Theory	✓		✓	✓	✓
		Table				✓	✓
	Challenges	Theory			✓		✓
		Table					✓
	Modality	Theory		✓	✓		✓
		Table		✓			✓
	Cross Domain	Theory			✓		✓
		Table					✓
	Metric Learning	Theory	✓		✓	✓	✓
		Table					✓

Table 2 Re-Id benchmark datasets

Dataset Type	Dataset	Release Year	Camera	Identity Count	Images/Videos Count	Challenges	Top Results—Rank 1 Accuracy (%)
Image Re-Id Datasets	VIPeR [34]	2007	2	632	1264 images	Viewpoint variation	67.21 [48]
	iLIDS [165]	2009	2	119	476 images	Illumination variation and occlusion	82.20 [5]
	GRID [76]	2010	8	1800	-	Viewpoint, lightning and color variations, occlusion	28.00 [84]
	CAVIAR4RE-ID [17]	2011	2	72	720 images	Resolution, light and pose variations, occlusion	53.60 [11]
	CUHK01 [58]	2012	2	971	1942 images	Pose, viewpoint and lightning variations	98.73 [86]
	RGBD-ID [7]	2012	1 (RGB-D)	79	316 images	View and clothing variation for same identities	76.70 [90]
	CUHK02 [56]	2013	10	1816	7264 images	Illumination and pose variations, partial occlusion	-
	CUHK03 [57]	2014	2	1360	13,164 images	Alignment variation, occlusion, missing body parts	96.43 [143]
	Market-1501 [164]	2015	6	1501	32,668 images	Illumination, scale and pose variations, partial occlusion	95.34 [86]
	Kinect-Re-Id [82]	2015	1 RGB-D	71	483 videos	Viewpoint and lightning variations	99.40 [90]
	DukeMTMC-Re-Id [33]	2017	8	1852	-	Illumination, view and pose variations, background clutter, occlusion	88.19 [86]
	RegDB [81]	2017	1 RGB 1 IR	412	4120 images 4120 thermal images	Pose, distance and lightning variation	48.43 [27]
	SYSU-MM01 [123]	2017	4 RGB 2 IR	491	287,628 images 15,792 IR images	Color and exposure variation	65.10 [26]
	Airport [49]	2019	6	9651	3.13 images per person on average	Viewpoint and illumination variation, detection error, occlusion, background clutter	-

Table 2 (continued)

Dataset Type	Dataset	Release Year	Camera	Identity Count	Images/Videos Count	Challenges	Top Results—Rank 1 Accuracy (%)
Video Re-Id Datasets	ETHZ [23]	2007	2			Partial occlusion	93.00 [18]
	PRID2011 [40]	2011	2	983	100 to 150 images per individual	Viewpoint, pose, lightning and background variations	98.80 [52]
	3DPES [4]	2011	6	200	500 videos		72.23 [169]
	iLIDS-VID [118]	2014	2	300	600 videos, 73 frames per video on average	Viewpoint and lightning variation, occlusion, cluttered background, similar clothing for different identities	88.00 [152]
	MARS [162]	2016	6	1261	20,715 videos	Viewpoint and pose variations, partial occlusion, detection error, small inter-class and large intra-class variation	87.30 [94]
	DukeMTMC-VideoRe-Id [136]	2019	-	1812	12 frames per second in a tracklet, 2196 training tracklets, 2636 testing tracklets	Lightning, pose and viewpoint variation, noisy background, occlusion	-

Re-Id datasets can be broadly classified into two types: image-based and video based datasets.

Table 3 Triplet loss improvements for Re-Identification

References	Year	Triplet Loss Improvements	Benefit
Ding et al. [20]	2015	Triplet Loss	Works on relative distance among intra-class and inter-class identities
Cheng et al. [16]	2016	Improved Triplet Loss	Ensures intra-class compactness
Hermans et al. [39]	2017	Batch Hard Triplet Loss	Removes triplet mining step overhead. Learning from similar inter-class and varying intra-class samples
Zhu et al. [173]	2017	Hash Based Triplet Loss	Ensures that the hamming distance of hash codes from the anchor and positive sample is less than that of anchor and negative sample
Su et al. [96]	2018	Attribute Triplet Loss	Learns a large number of human attributes considering their contextual cues
Wu et al. [128]	2019	OIM + Improved Triplet Loss	Learn similarity metric and fully utilize label information of samples
Yuan et al. [147]	2019	Mini-Cluster Loss	Reduce intra-class and enlarge inter-class differences
Si et al. [93]	2019	Compact Triplet Loss	Reduce intra-class and enlarge inter-class differences
Yang et al. [144]	2019	Adaptive Nearest Neighbour Loss (ANN)	Solves slow convergence and local optima for triplet based loss
Choe et al. [18]	2019	Mixed Distance Maximization Loss	Maximize intra-distance and keep triplet distance larger than sample distance
Zhou et al. [170]	2019	Symmetric Triplet Loss	Gradients derived for positive samples are symmetric allowing consistent minimization of intra-class distances
Fan et al. [24]	2019	Sphere Softmax Loss	Maps image samples to a hypersphere manifold
Zhang et al. [160]	2020	Hybrid Triplet Loss (HTL)	Learns domain invariant and camera invariant properties
Jiang et al. [47]	2020	Weighted Triple-Sequence Loss (WTSL)	Reduce impact of outlier frames in video sequences
Zhang et al. [155]	2020	Wasserstein Triplet Loss	Uses the Wasserstein Distrance to rearrange global distance between samples
Sikdar et al. [92]	2020	Batch Adaptive Triplet Loss	Exponential learning from hard positives compared easier positives in triplet scheme

Table 4 Contributions of attention modules in Re-Id methods

Reference	Year	Region of Attention	Benefit
Yang et al. [143]	2019	Whole body and body parts	Discriminative feature extraction
Bao et al. [6]	2019	Body parts	Robust to part misalignment and background clutter
Zhou et al. [170]	2019	Image foreground	Robust to background clutter
Wu et al. [129]	2019	Spatial regions in video frames, temporal pooling over entire video	Extract discriminative Re-Id features from essential frames within video
Li et al. [54]	2019	Convolutional features	Learns interdependence of channels within convolutional features
Zhang et al. [152]	2019	Video frames	Select informative frames for dimensionality reduction of features
Wan et al. [106]	2019	Spatial region of input images	Local parts discovery
Tay et al. [102]	2019	Physical appearance attributes like gender, hair, upper clothing color, lower clothing color, pant etc	Discriminative representation of identity based on appearance attributes (Fig. 3)
Hou et al. [42]	2019	Generated Frames created by GAN generator	Generate occluded regions in video frames using temporal attention module to remove occlusion
Subramaniam et al. [97]	2019	Common visual features across frames of a video	Robust to background noise and extracts common features in video frames
Chen et al. [10]	2019	Spatial and Channel-wise attention	Improved attention quality due to self-critical attention learning
Zhang et al. [154]	2020	Video frames focussed at multiple scales	Spatio-temporal video representations
Li et al. [60]	2020	Global image features at multiple scales	Robust to spatial misalignment and local feature dependencies
Qian et al. [86]	2020	Spatial regions of input, multi-scale features	Discriminative feature extraction
Zhang et al. [155]	2020	Global and local features	Robust to misalignment, helps to distinguish between important and misleading parts

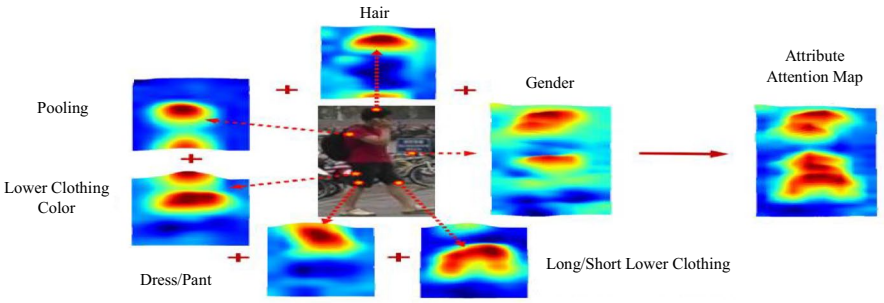


Fig. 3 Attribute Attention Map (AAM) generated from six heat maps corresponding to attributes such as gender, hair, clothing etc. [102]

2.2 Review methodology

This review includes journal papers, conference and workshop papers from several well-known repositories including IEEE Xplore, ScienceDirect, Springer, ACM and Google Scholar. The keywords used to search for relevant contributions include "person re-identification", "Re-Id", "deep", "deep learning", "review" and "survey". Due to the growing popularity of deep approaches, this initial search resulted in a comprehensive list of contributions. Higher priority was given to publications from high-quality journals such as IEEE Transactions, Pattern Recognition, Neurocomputing etc. and top conferences such as CVPR, ECCV and ICCV. A separate search was conducted to include Re-Id dataset contributions. Finally, the included papers were analyzed and a taxonomy of deep Re-Id methods was formulated as demonstrated in Fig. 2. Figure 4 shows the graph of year-wise deep Re-Id contributions cited in this review clearly depicting a high percentage of recent deep Re-Id works.

2.3 Reseach questions

This section presents the research questions explored in this survey. Specifically, it highlights the key issues of person re-identification and how different research contributions aim to address them. Exploring several research papers on person re-identification, various key ideas have been identified. *Firstly,* deep learning-based approaches are primarily about designing novel architectures to solve a given problem. Hence, understanding the different types of deep architectures for person re-identification is important. *Secondly,* researchers face several challenges in person re-identification and identifying these barriers is crucial. *Thirdly,* loss functions can be modified to enforce robust feature learning capability. *Fourthly,* how can the recently proposed attention mechanism be used for person re-identification? cation? *Finally,* how is it possible to build multi-modal approaches for person re-identification systems?

In this regard, the key research questions for person re-identification are as follows:

- *What are the different types of deep learning architectures that are used for person re-identification?*
- *What are the major challenges for person re-identification systems?*

Table 5 Re-Id contributions based on different architectures

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Ding et al. [20]	Triplet	Propose "Triplet loss" the utilizes three samples to learn discriminative features and also propose a triplet generation scheme	Triplet loss	Proposed model reduces the intra-class variations and increases inter-class variations via L2 distance metric based triplet inputs	Requires additional preprocessing step to formulate triplets with positive, negative and neutral samples
Zhu et al. [172]	Classification	Combine Softmax and Center Loss [121] to improve discriminative capability of CNN features	Softmax + Center Loss	Reduces the intra-class variations via a combined combination of softmax and center loss	Requires large number of training samples per identity
Zhu et al. [173]	Triplet + Part	Propose a "Part Based Deep Hashing" (PDH) network that integrates hashing for higher efficiency of large scale Re-Id. Implement a triplet loss that reduces the Hamming distance of positive sample parts and increase that of negative sample parts	Triplet loss	Proposed model utilizes hasing for fast approximate nearest neighbour search	Requires large-scale reid datasets to utilize the full strength of hashing mechanism
Huang et al. [45]	Part	Propose a novel method DeepDiff to evaluate similarity between corresponding parts using original data, feature maps and spatial variations from three deep subnets	Softmax loss	Utilizes a novel pyramid partition architecture for difference granularities of human parts	Final score is attained by accumulating similarity scores of individual parts
Koo et al. [53]	Part	Use information from face and body to obtain discriminative representations in indoor camera surveillance environment	Softmax loss	Proposed a multimodal human recognition method that used both face and body	Proposed approach uses multiple CNNs which increases the computational cost of training the model

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Tao et al. [101]	Triplet	Propose a "Deep Multi-View Feature Learning" (DMVFL) method that fuses handcrafted and deeply learned features to obtain robust representations	Triplet loss	Proposed approach extracts complementary information from deep features and cross-view quadratic discriminant analysis metric learning	Proposed model is small and has limited learning capability
Su et al. [96]	Triplet	Propose a three-stage "Weakly Supervised Multi-Type Attribute Learning Framework" using a novel "Attribute Triplet Loss" to predict visually invariant features containing contextual cues	Attribute triplet loss	Extract human attribute features that are invariant to changes in pose and viewpoint	Proposed model does not consider the spatial locations of human attributes
Zhang et al. [159]	Classification + Verification + Part	Learn deep features from body parts and entire body using verification and classification models respectively to obtain final representation	Softmax loss	Proposed model combines the benefits of part-based and body-based feature learning	Proposed method requires the additional processing of pairing positive and negative samples
Zhong et al. [168]	Classification	Propose a novel "Inter-Center Loss" to improve the discriminative capability of CNN features	Softmax + Center + Inter-Center Loss	Combining multiple losses helps to reduce intra-class variations	Combining three losses requires the additional tuning of loss weight parameters

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Fan et al. [24]	Classification	Propose a novel CNN based network "SphereRe-Id" adopting a novel sphere loss mapping sample images to hypersphere manifold and a balanced sampling strategy to address class imbalance	Sphere Softmax Loss	Proposed a novel loss function that lears the underlying feature embedding hypersphere which is often ignored methods	Sphere softmax is computationally expensive due to the calculation of cosine similarity between two vectors
Bao et al. [6]	Classification + Attention	Propose a dual-branch CNN based network having a global branch to process features from overall human body and an attention branch to selectively focus on attentional part of information from input	Softmax Loss	Uses visual attention mechanism to focus on spatial regions of important body parts	Basic CNNs are used as individual branches. More recent architectures will extract better features
Yang et al. [143]	Part based + Attention	Introduce a novel attention driven multi-branch network to learn discriminative representations from whole body and body parts focussing on spatial and channel-wise information	Softmax loss	Uses attention mechanism to focus on important spatial and channel regions	Semantic part parsing was not used in the proposed network
Yan et al. [142]	Part + Attention	Propose a feature attention block for Re-Id that pays attention to sliced part features in a weighted manner to highlight the most discriminative part regions which are robust to misalignment	Softmax loss	Uses attention mechanism to learn weighted part-level features that are discriminative	Sample re-weighting is not used in the proposed approach

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Zhou et al. [170]	Triplet + Part + Attention	Propose a novel "Foreground Attentive Neural Network" (FANN) that utilizes a foreground attentive subnetwork and a novel regression loss function to learn foreground regions which is then fed to body part features using novel symmetric triplet loss	Regression loss + Symmetric Triplet loss	Proposed model contains a foregroud mask reconstruction mechanism to suppress irrelevant background information	Computing multiple losses increases the computational overheads
Wu et al. [126]	Classification + Triplet	Combine classification loss, triplet loss and center loss to constrain Euclidean distance of same identity samples closer than those of different identities	Softmax loss + Triplet loss + Center loss	Proposed model enjoys the benefits two types of achitectures—identification and triplet model	Computing multiple losses increases the computational overheads
Zhang et al. [158]	Classification + Triplet	Introduce a dual-branch "Multi Branch Slice-Based Network" (MSN) learning multi-level local and global features using a novel "triplet-center loss"	Triplet-Center loss	Proposed model extracts multi-level pyramid features from global region and local body part regions	Dual-branch architecture makes the proposed approach computationally expensive
Zhao et al. [161]	Triplet	Introduce a multi-level triplet model MT-net, extracting multi-level features which are both global and local	Triplet loss	Proposed model combines fine features from shallow layers and coarse features from deep layers	The three branch architecture contains GoogleNet in each branch which makes it computationally heavy

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Yuan et al. [148]	Classification + Triplet	Introduce a deep joint embedding learning framework that uses classification and an improved triplet loss. The improved triplet loss works on hard triplets generated	Softmax loss + Improved Triplet loss	Proposes an improved triplet loss that swaps the neutral and positive sample in a triplet to compute robust loss	Performance on large-scale reid datsets is unknown
Wu et al. [128]	Classification + Triplet	Combine classification and triplet loss to make full use of labels as well as learn similarity measure simultaneously	Online Instance Matching loss + Triplet loss	Use of online instance matching loss is better for larger datasets as compared to Softmax loss	Local descriptor features are not extracted in the proposed approach
Tian et al. [103]	Part	Propose a two-branch CNN to combine learning global and local part level features simultaneously	Softmax loss	Weighted combination of local and global losses allows for intelligent local and global feature extraction	Dual-branch architecture is computationally heavy
Li et al. [54]	Classification + Verification + Attention	Propose a novel network with attention module to highlight essential features and a multi-loss function to reduce intra-class distance and increase inter-class distance	Cross-entropy loss + novel verification loss	Attention mechanism highlights important regions of features and multi-loss function helps in faster model convergence	Siamese networks contains two CNNs adding a large number of parameters to the network

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Yuan et al. [147]	Triplet	Propose a novel "mini-cluster" loss that ensures the largest distance of same identity samples (inner divergence) to be less than the smallest distance of different identity samples (outer divergence)	Mini-cluster loss	Proposed mini-cluster loss reduces intra-class variations and expands inter-class differences	The proposed approach does not extract effective detailed local features
Ling et al. [68]	Classification + Verification	Introduce MTNet with four losses for identification and verification of person identity and person attributes	Softmax loss	Attribute loss helps to extract discriminatory pedestrian information	The proposed approach does not utilize complementary benefits of attributes and reid tasks
Si et al. [93]	Classification + Triplet	Propose a novel "Compact Triplet Loss" that improves the batch hard triplet loss to reduce intra-class variations and increase inter-class differences. Combine with a classification loss for better discrimination capability	Compact triplet loss + Softmax loss	Proposed loss is robust against pose variations	Limited generalization to unseen dataset examples
Tian et al. [104]	Verification + Triplet	Propose a novel "Adaptive Verification Loss" (ADV loss) that learns only from meaningful hard sample pairs mined by using a weighted triplet loss	ADV loss + Triplet loss	Proposed loss handles unbalanced datasets better than other commonly used methods	Results are highly prone to distortion due to outlier examples

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Zhong et al. [167]	Classification + Verification	Introduce a novel "Feature Aggregation network" (FAN) network to learn features from various layers of deep network along with multiple losses	Softmax loss	Multiple feature extraction techniques and multi-loss learning approach for discriminative learning	High forward pass time per batch
Wang et al. [119]	Part	Propose a novel "Part-Based Pyramid Loss" that takes quadruplet input samples and learns body part features using relationship of distance and angle among samples	Part based pyramid loss	Proposed method captures both local and global features	Part-based pyramid loss requires large amount of training data and hence, is ineffective for small datasets
Yao et al. [145]	Classification + Part	Propose a "Part Loss Network" (PL-net) that trains on body part and global features	Classification loss + Part loss	Robust to occlusion and viewpoint variants	Part-based loss function increases the training time of the model
He et al. [38]	Classification + Verification	Adopt the "lifted structured loss" due to its superiority to contrastive and triplet losses. Combine it with identification loss to learn relative identity information from pairs and true identity information	Softmax loss + Lifted structured loss	Proposed model minimizes intra-class variations and maximizes inter-class variations by using lifted structured loss	Using lifted structured loss limits model scalability as the computational overhead increases with number of samples and feature dimension

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Choe et al. [18]	Triplet	Consider intra-distance between positive samples of a triplet and distance between triplets using a novel "mixed distance" function to improve Re-Id performance	Mixed distance loss	Proposed method solves the local optima problem by using "mixed distance maximization "	Limited generalizability to unseen data
Quispe et al. [87]	Classification + Triplet	Propose a novel "Saliency Semantic Parsing Re-Id" (SSP-Re-Id) network. The Saliency guided subnetwork learns from essential parts of image while semantic parsing guided subnetwork deals with Re-Id challenges	Softmax loss + Triplet loss	Proposed model is robust to pose / illumination variations and occlusion	High computational overheads due to saliency and semantic parsing with deep learning networks
Wan et al. [106]	Part + Attention	Propose a novel "Concentrated SPR network" (CSPR-Net) having a "constrained attention module" to find discriminative local parts that work better than body parts and a novel "statistical-positional-relational (SPR) descriptor that gives better the performance than global features	Classification loss	Proposed model is robust to pose and illumination variations	Proposed model is not robust to occlusion

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Zhang et al. [155]	Classification + Triplet + Attention	Propose a novel triplet loss based on the "Wasserstein distance" (Earth Mover distance) to handle the part misalignment problem using part probabilities obtained from attention maps and part features	Softmax loss + Wasserstein triplet loss	Proposed method is view-invariant and lighting invariant	The proposed approach has limited scalability to larger datasets because computing Wasserstein distance between two images is computationally expensive
Li et al. [55]	Part	Propose a novel "Attributes-Aided Part Detection and Refinement Network" (APDR) that uses attribute learning for part localization handling the misalignment problem. Attribute features are fused to obtain discriminative features	Softmax loss + Triplet loss	Proposed model is robust to occlusion	Proposed model relies on extracting attribute information which may not always be consistent or reliable
Qian et al. [86]	Classification + Triplet + Attention	Propose a novel "Multi-Scale Deep Architecture" (MuDeep) having a "multi-scale deep learning layer" to learn features at different scales and a "leader-based attention learning layer" to determine optimal weighting for features from each scale	Softmax + Triplet loss	Attention mechanism of the proposed model dynamically finds the spatial regions of importance at each feature scale	Increased computational cost due to multiple attentional modules and feature fusion steps

Table 5 (continued)

Ref	Architecture	Key idea	Loss Function	Pros	Cons
Bai et al. [3]	Classification + Triplet + Part	Propose a three-branch "Deep Person" framework that learns contextual body part information using LSTM and learn discriminative features using identification (global and part level) and triplet losses	Softmax loss + Triplet loss	Proposed approach considers human body as a sequence of body parts and learns their spatial context using LSTM	Proposed approach does not use attention mechanism
Zhang et al. [150]	Part	Propose a dual-branch "Heterogeneous Part-Based Deep Network" (HPDN) to learn part based features using batch hard triplet and cross entropy loss	Batch hard triplet loss + Cross-entropy loss	Model searches for the person based on the local part alignment score	The proposed dual-branch architecture is computationally heavier than single branch models
Li et al. [60]	Classification + Triplet + Attention	Introduce a self-attention guided model that learns weighted features from different regions of human image	Softmax loss + Triplet loss	Self-attention module extracts weighted person regions and non-local features	Multi-scale feature fusion increases the model's computational complexity

Table 6 Contributions robust to Re-Id challenges: variations in Pose (P), Scale (S), Lightning (L) and View (V), Occlusion (O), Background clutter (B) and Misalignment (M)

Reference	Year	Key Idea to Avert Re-Id Challenges	Robust to Challenges						
			P	S	L	O	B	V	M
Yu et al. [146]	2018	Use skeleton joint information and cloth-color type features to achieve pose and lightning invariance	✓		✓				
Feng et al. [26]	2018	Use Cross-View Euclidean constraint (CV-EC) to reduce distance of deep features of an identity from multiple views						✓	
Zhou et al. [169]	2018	Propose a "self-paced learning" (SPL) method to isolate noisy samples in weighted way using model age and iteration and train CNN model with faithful samples gradually moving from easy to hard samples				✓	✓		
Chen et al. [12]	2018	Fuse attribute features learned from part-specific CNN and fuse them with Local Maximal Occurance (LOMO) features to obtain robust features	✓		✓		✓	✓	
Fu et al. [28]	2019	Introduce a two stream spatial segmentation network that derives spatial and fine local features	✓		✓			✓	
Zheng et al. [166]	2019	Use a novel "alignment branch" to learn the affine transformation of high level convolutional features utilizing a spatial transformer network that crops images with too much background and pads zero borders to missing part features thereby solving scale variation and misalignment		✓			✓		✓
Chen et al. [14]	2019	Study correlation among cross-view visual data from multiple camera views to compose view-specific representations						✓	
Luo et al. [78]	2019	Propose a novel "Dynamically Matching Local Information" (DMLI) method that automatically aligns horizontal stripes from input samples without any labelling supervision or pose estimation	✓						
Zhou et al. [170]	2019	Use a novel "foreground attentive subnetwork" having a decoder trained on a novel "local regression loss" to create a binary mask suppressing background					✓		
Qi et al. [85]	2019	Train CNN on six Re-Id datasets to learn robust features			✓			✓	
Yang et al. [144]	2019	Learn spatial dependencies among local regions of pedestrians in both horizontal and vertical directions using LSTM to overcome occlusion				✓			
Zheng et al. [163]	2019	Propose a novel "pose invariant embedding" (PIE) by constructing a novel PoseBox through pose estimation and then training using two-stream PoseBox Fusion network	✓						
Wei et al. [122]	2019	Estimate four human key points that are robust to pose and view variations. Head, upper body and lower body features are generated using these key points and a four-stream CNN is trained to generate the novel "Global Local Alignment Descriptor" (GLAD) features from both global and local regions	✓						✓
Wu et al. [133]	2020	Use adversarial learning to learn asymmetric transformations to transform view-specific distribution to a generic view invariant feature space						✓	
Li et al. [55]	2020	Use attribute learning to detect body parts thereby solving part misalignment problem							✓

Table 6 (continued)

Reference	Year	Key Idea to Avert Re-Id Challenges	Robust to Challenges						
			P	S	L	O	B	V	M
Tang et al. [100]	2020	Propose a novel "Gradual Background Suppression" method that extract CNN features based on different weights assigned to body parts and background thereby suppressing the background					✓		
Sikdar et al. [92]	2020	Resize input to different scales and convolve with a fixed size filter to obtain multi-resolution pyramid which is re-scaled back to fixed size to obtain scale invariant features		✓					
Li et al. [60]	2020	Propose a "multi-scale attention" model evaluating important person regions in a weighted fashion and train on features fused globally and locally using cross-entropy and triplet loss	✓			✓		✓	

Table 7 Image based Re-Id contributions

Reference	Year	Key Idea	Benefit	Dataset	Rank 1 Result (%)
Wu et al. [124]	2018	Introduced a deep embedding approach using optimized robust features, positive mining and local adaptive similarity learning	Discriminative features	VIPeR	49.04
				CUHK01	71.60
				CUHK03	73.02
				Market-1501	84.14
Wu et al. [125]	2018	Introduce a multiplicative integration gating function combined with Hamadard product	Cross-view feature alignment	VIPeR	49.11
				CUHK03	73.23
				Market-1501	67.15
				Market-1501	93.91
				DukeMTMC	83.35
Ke et al. [50]	2018	Introduce ID-AdaptNet to adapt “seen” identity features to “unseen” identities	Discriminative features	CUHK03	30.36
				Market-1501	81.59
				DukeMTMC	67.77
Zhang et al. [151]	2019	Use group symmetry theory to extract and utilize information from middle layers of Resnet50 (ResGroupNet)	Discriminative features	CUHK03	71.20
				Market-1501	92.80
				DukeMTMC	86.20
				Market-1501	93.50
				DukeMTMC-Re-Id	85.00
Wu et al. [130]	2019	Introduce a five-branch deep model that learns body features in horizontal and vertical directions, relationship between feature channels and part features	Discriminative features	CUHK03	95.00
				Market-1501	94.70
				DukeMTMC-Re-Id	86.70
Liu et al. [73]	2019	Fuse Gaussian features with deep features	Fused discriminative features	Market-1501	84.40
				VIPeR	57.20
Zhang et al. [153]	2019	Introduce a PAAN network using layered partition strategy to fully utilize part-level and global attributes	Fused discriminative features	Market-1501	91.86
				DukeMTMC-Re-Id	81.73
Tian et al. [103]	2019	Introduce a two-branch RJLN network to jointly learn global and local features	Fused discriminative features	CUHK03	66.60
				Market-1501	93.70
				DukeMTMC-Re-Id	85.50

Table 7 (continued)

Reference	Year	Key Idea	Benefit	Dataset	Rank 1 Result (%)
Wu et al. [131]	2019	Introduce a multi-branch MFML network to represent features from multiple layers	Weighted multi-layered fused features	CUHK03	94.40
				Market-1501	92.50
				DukeMTMC-Re-Id	84.00
Zhao et al. [161]	2019	Introduce a multi-level triplet model MT-net, extracting multi-level features which are both global and local	Fused discriminative features	CUHK03	79.34
				Market-1501	81.95
Wang et al. [113]	2020	Propose a novel exclusively regularized softmax objective function	Multi-scale multi-patch features	CUHK03	70.40
				Market-1501	93.70
				DukeMTMC	84.40
Wang et al. [114]	2020	Fuse handcrafted features from local and global regions with deep features	Fused discriminative features	VIPeR	52.22
				CUHK01	71.91

Table 8 Cross modality Re-Id contributions

Reference	Year	Modality	Key Idea	Device
Ren et al. [89]	2017	RGB + Depth	Fuse anthropometric features from depth images and visual features from RGB images using a novel "multi-modal fusion layer" to obtain discriminative features	RGB Camera + Kinect V1 sensor
Feng et al. [27]	2019	RGB + Infrared Data	Extract "modality specific representations" (MSR) from modality-specific networks trained on a modality-specific loss to learn discriminative features from each domain. Use a cross-modality Euclidean constraint to learn modality-invariant features	RGB + Infrared Cameras
Xiang et al. [137]	2019	RGB + Infrared	Propose a dual-branch neural network that fuses modality specific information from two branches using multiple granularity network (MGN) to obtain modality shared features	RGB + Infrared Cameras
Ren et al. [37]	2019	RGB + Depth	Propose a novel "uniform and vibrational deep learning" (UVDL) method that uses an auto-encoder to visible and deep features extracted from two neural nets into a common features space	RGB-D camera (Kinect)
Chang et al. [9]	2020	RGB + Textual	Combine visual CNN features with textual CNN features from textual descriptions including details of gender, clothes color and bag type information to obtain a generalized feature embedding	RGB Camera
Gohar et al. [32]	2020	Gait Data	Propose a novel "non-visual gait based" Re-Id method that uses gait data captured from accelerometer and gyroscope to learn discriminative features using Gated Recurrent Unit (GRU) to capture discriminative temporal information from input sequences	Accelerometer, Gyroscope
Wang et al. [117]	2020	RGB + Infrared	Propose a novel "multi-patch networking network" (MPMN) that utilizes a single deep neural net to process both RGB and thermal images. A novel "multi-patch modality alignment" loss mines for hard subspaces. A novel "cross-patch correlation distillation" (CPCD) loss enforces cross-patch similarity to boost cross-modality embedding. A novel "patch aware priority attention" (PAPA) method prioritizes training of hard patch tasks over others	RGB + Infrared Cameras

Table 9 Cross-domain Re-Id contributions

Reference	Year	Cross-Domain Approach
Xiao et al. [138]	2016	Learn generic CNN features embedding from multiple dataset domains and use a novel "Domain Guided Dropout" to mute neurons learning domain specific information thereby improving Re-Id performance
Ganin et al. [29]	2016	Use few standard layers and a novel "gradient reversal layer" to learn from labelled source domain samples and unlabelled target domain samples
Xu et al. [141]	2019	Propose a novel "Deep Augmented Attribute Network" (DAAN) to learn augmented feature representations using augmented features and labels from source dataset and refine the learned features to unlabelled target dataset
Zhou et al. [171]	2019	Propose a novel "Multi-Camera GAN" (CTGAN) to transfer source dataset samples to multi-camera styles of target dataset
Genc et al. [30]	2019	Perform domain adaptation by training on different dataset combinations, learning part specific features and learning features form multiple layers and use a CycleGAN to perform camera view adaptations
Wang et al. [109]	2020	Propose a novel "Deep Multi-Task Transfer Network" (DMTNet) network for unsupervised cross domain Re-Id including cluster number estimation algorithm, learning of attribute attention importance and transfer of specific multi-task learning across domains
Sun et al. [99]	2020	Propose a novel "Conditional Transfer Network" (cTransNet) implementing conditional viewpoint transfer using StarGAN and obtain hybrid feature embeddings from original and transformed images to obtain similarity rankings
Zhang et al. [160]	2020	Propose a novel "Dual Generation Learning" (DGL) method to transfer source dataset images to target style domain and target dataset images to source camera styles to obtain better Re-Id results

Table 10 Deep metric learning methods

Reference	Year	Metric Learning Method	Short Form	Key Idea
Hu et al. [43]	2016	Deep Transfer Metric Learning	DTML	Cross domain metric learning
Lin et al. [67]	2017	Generic Similarity Metric	-	Robust to translation and shearing deformation
Zhu et al. [174]	2018	Deep Hybrid Similarity Learning	DHSL	More discriminative than Euclidean or Cosine distance based similarity
Duan et al. [22]	2018	Deep Localized Metric Learning	DLML	Metric learning over locally varying data
Hu et al. [44]	2018	Sharable and Individual Multi-View Deep Metric Learning	MvDML	View invariant metric learning
Chen et al. [15]	2018	Pose Invariant Deep Metric Learning	PIDML	Pose invariant metric learning
Ren et al. [88]	2019	Deep Structured Metric Learning	-	Robust person re-identification
Ding et al. [21]	2019	Robust Discriminative Metric Learning	RDML	Robust to noise
Xiong et al. [139]	2019	Multiple Deep Metric Learning	-	Instead of feature extraction, utilize a stacked auto-encoder to recognize individuals using multiple similarity probabilities from Softmax regression models

Table 11 Video based deep Re-Id contributions

Reference	Year	Key Idea	Dataset	Rank 1 Result (%)
Sun et al. [98]	2018	Extract visual features using a "Two-Branch Appearance Feature sub-structure" (TAF) and temporal features using a "Optical Flow Temporal Feature sub-structure" (OTF) and use a pair of Siamese networks to learn similarity measure among pairwise visual and temporal features. A saliency learning fusion layer learns fusion of global and local appearances	MARS iLID-VIDS PRID2011	73.00 59.00 79.00
Wang et al. [108]	2018	Perform image-to-video Re-Id using a novel end-to-end "Point-to-Set Network" (P2SNet) that takes both image and video as input and jointly learns their features using point-to-set distance metric. A *k*NN-triplet module helps to focus only on relevant video frames	MARS iLID-VIDS PRID2011	55.25 73.31 40.00
Meng et al. [80]	2019	Learn view-specific and feature-specific transformations using a novel "Deep Asymmetric Metric Learning" from a two stream neural net to counter view-specific bias and feature-specific bias in appearance and motion features due to variations in view, lightning, background clutter etc	MARS iLID-VIDS PRID2011	74.65 77.33 87.00
Wu et al. [129]	2019	Propose a novel "Siamese attention architecture" to jointly learn feature representation and similarity measure for video Re-Id using an attention mechanism focussing on relevant frames (when) and regions of interest within frames (where)	MARS iLID-VIDS PRID2011	86.60 86.50 98.80
Ksibi et al. [52]	2019	Propose a "Deep Spatio-Temporal Appearance" (DSTA) descriptor that uses a "Deep Salient-Gaussian Wighted Fisher Vector" (SGFV) to exploit trajectory information to handle misalignment of person tracklets and eliminate background noise using Gaussian and Saliency maps	MARS iLID-VIDS PRID2011	76.70 80.00 92.70
Zhang et al. [152]	2019	Propose a novel "Self-and-Collaborative Attention Network" (SCAN) that utilizes two attention sub-networks (self-attention subnetwork and collaborative attention subnetwork) to select features from informative video frames and align discriminative frames from probe and gallery frames respectively and finally use a generalized similarity measure to compare video pair representations	MARS iLID-VIDS PRID2011	87.20 88.00 95.30
Wu et al. [132]	2019	Utilize novel "3D Person VLAD Aggregation layer" based on the vector of locally aggregated descriptors to capture motion based features along with appearances which is usually missed by 2D ConvNets and learn "global representations" for a full length video robust to occlusion and misalignment using soft attention module to learn 3D part alignment	MARS iLID-VIDS PRID2011	80.80 69.40 87.60
Liu et al. [71]	2019	Propose a novel "Dense 3D Convolutional Network" (D3DNet) that uses numerous three dimensional dense blocks and transition layers to extract discriminative features from appearance and temporal domains capturing visual and motion cues (both short term and long term) from videos without additional modules. Implement a combination of identification and center loss to reduce intra-class disparity and increase inter-class disparity	MARS iLID-VIDS	76.00 65.40

Table 11 (continued)

Reference	Year	Key Idea	Dataset	Rank 1 Result (%)
McLaughlin et al. [79]	2019	Extract frame-wise CNN features and use a temporal pooling recurrent layer to combine all time step features and obtain feature representation for entire video sequence	iLID-VIDS	58.00
			PRID2011	70.00
Zhang et al. [154]	2020	Propose a novel "Multi-Scale Spatial–Temporal Attention" (MSTA) model that pays attention to different regions within each video frame at different scales to incorporate essential regions into whole video spatio-temporal representations. MSTA contains an encoder to extract frame wise features and an aggregator to fuse features	MARS	82.28
			iLID-VIDS	70.10
			PRID2011	91.20
Wu et al. [134]	2020	Use variational recurrent neural networks (VRNNs) to conduct a deep few-shot adversarial learning to extract discriminative features that are view invariant	MARS	54.60
			iLID-VIDS	60.10
			PRID2011	79.20
Avola et al. [2]	2020	Propose a novel "LSTM based Re-Id Hashing model" that exploits bone proportion, gait and movement features of 2D skeletons extracted from RGB video frames. LSTM is used to learn temporal correlation between different frames while two dense layers are responsible for implementing bodyprint hashing via binary coding capable of unbounded labelling of all individuals possible	MARS	86.50
			iLID-VIDS	73.40
			PRID2011	82.70
Wu et al. [135]	2020	Propose a novel "Adaptive Graph Representation Learning" method that uses adaptive structure-aware adjacency graphs via graph neural networks (GNN) highlighting two kinds of relations: pose alignment connection to capture human part relations and feature affinity connection to model semantic relationship among features from various regions across frames. A novel regularization is used to capture temporal resolution invariant features for the entire video sequence	MARS	89.80
			iLID-VIDS	83.70
			PRID2011	93.10
			Duke-MTMC VideoRe-Id	96.70
Jiang et al. [47]	2020	Propose a novel framework "Spatial Transformed Partial Network" (STPN) that aligns frames to extract robust regional features. A novel "Weighted Triple-Sequence Loss" (WTSL) is used to exclude outlier frames in video-level features. Model is jointly optimized for frame-level and video level features	MARS	85.90
			iLID-VIDS	82.20
			PRID2011	95.20

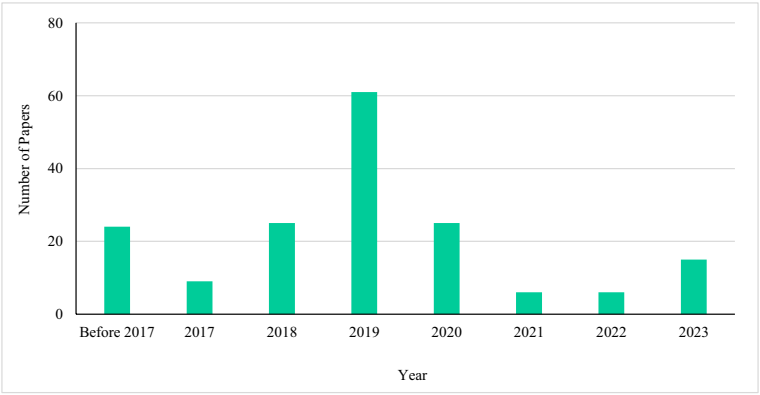


Fig. 4 Count of cited articles (year-wise)

- *What novel loss functions have been proposed to compute robust features towards person re-identification?*
- *How can the attention mechanism boost the performance of person re-identification systems?*
- *Can multi-modal person re-identification systems perform better than methods with single modality?*

This survey addresses these research questions. Section 4.1 talks about different types of architectures for person re-identification. Section 4.2 gives the major challenges. Section 4.1.3 presents several novel losses for person re-identification. Section 4.1.5 presents the attention-based person reid methods. Multi-modal person re-identification research has been discussed in Section 4.3

2.4 Comparison with existing reviews

While the number of deep Re-Id implementations have grown exponentially in recent years, there are very few reviews/surveys keeping pace with the growth of deep Re-Id research. This section presents a comparative analysis of this review against some existing deep learning based Re-Id surveys. Compared to other surveys, this review includes the latest research publications (up to 2020). Table 1 presents Re-Id comparisons on the basis of several deep learning aspects such as architecture types, Re-Id datasets, challenges, modality, cross-domain approaches and metric learning methods. The green rows indicate the presence of theoretical analysis while the orange rows indicate the presence of tabular information.

Table 1 clearly shows the comprehensive nature of this review when compared to other existing surveys, presenting both theoretical and tabular analysis of various deep Re-Id aspects.

3 Benchmark datasets for person Re-Identification

Several benchmark datasets have been collected over the years to train and test the robustness of person re-identification systems. These datasets are useful in validating different re-identification approaches in terms of recognition accuracy. Table 2 gives detailed information about various re-id datasets.

3.1 Image based Re-Id datasets

The VIPeR [34] dataset has been created for viewpoint invariant pedestrian recognition. It contains 1264 images of 632 identities that have been captured from 2 cameras. The iLIDS [165] dataset has been acquired at an airport arrival hall and has 476 images of 19 identities from 2 cameras. The GRID [76] dataset is acquired from a busy underground train station having images of 800 identities from 8 cameras. CUHK03 [57] is another large scale re-id dataset having 13,164 images of 1360 identities collected using 2 cameras.

Datasets like Market-1501 [164], DukeMTMC-Re-Id [33] and CUHK02 [56] have used 6,8 and 10 cameras respectively, thereby increasing the number of camera views used to collect people images. A few multi-modal re-id datasets also exist such as the RGBD-ID [7], KinectRe-Id [82] based on RGB-depth images and RegDB [81], SYSU-MM01 [123] containing both RGB and infrared images. The largest image based re-id dataset is the Airport [49] collected from 6 cameras of a mid-sized airport containing images for 9651 identities. Figure 5 shows some sample images from the Market-1501 dataset.

3.2 Video based Re-Id datasets

PRID2011 [40] dataset contains 983 identities collected from 2 cameras. iLIDS-VID [118] contains 600 videos of 300 people from 2 cameras. Another popular video based re-id dataset is the large scale dataset MARS [162] which contains 20,715 videos of 1261 identities from 6 cameras. The most recent addition to the video re-id dataset is the DukeMTMC-VideoRe-Id [136] containing 1812 identities.

Considering the growth of re-id datasets over the years, several inferences can be made. *Firstly*, Re-Id datasets have grown both in number and scale which is a great benefit since deep models require large amount of samples for effective training. *Secondly*, the variety of samples within these datasets present numerous re-id challenges such as variations in pose, lightning and scale, occlusion, background clutter, same people wearing different clothing (large intra-class disparity) or different people wearing similar clothing (small inter-class disparity), thereby allowing deep models to learn

Fig. 5 Sample Images from Market-1501 dataset

effective generalization of appearance. *Thirdly*, very few datasets are multi-modal [7, 82] leading to an over reliance on RGB image and video based approaches. *Fourthly,* in a supervised environment, deep models require labelled samples for learning. As the size of datasets grow, it becomes less feasible to annotate them manually. While most of the datasets discussed above have manually annotated samples, datasets like Market-1501 or CUHK03 have used Deformable Part Model (DPM) [25] for sample labelling.

4 Image based deep Re-Id contributions

This section details the recent image based deep Re-Id contributions. These contributions have been categorized according to the following aspects: 1) Architecture types for Re-Id 2) Re-Id challenges 3) Data Modality for Re-Id 4) Cross-Domain Re-Id 5) Metric Learning for Re-Id. These categories are not exclusive and often overlap in various implementations but each has a distinct conceptual aspect to it.

4.1 Deep Re-Id architecture types

This section talks about the different kinds of architectures used for deep learning based re-identification. Specifically, the Re-Id contributions have been categorized as 1) Classification Models 2) Verification Models 3) Triplet Based Models 4) Part-Based Models 5) Attention Based Models as shown in Fig. 6.

4.1.1 Classification models for Re-Id

Classification models (also known as Identification models) consider Re-Id as a multi-class classification problem [62, 168]. Given a dataset with a finite number of identities and each identity having a number of samples, these models are trained using identity labels from samples. Classification models can be formally described as follows. Let there be an image gallery of κ people $P = \{p_1, p_2, p_3 ... p_\kappa\}$ with identity labels $L = \{l_1, l_2, l_3 ... l_\kappa\}$. The model is trained with labels L from various sample identities of people P. After training, given a query sample p_x having identity l_x the classification model tries to output a high score for label x and low score for all other identity labels. Since Re-Id models are trained on samples from various identities, they require large number of samples per identity to capture diverse features from each individual. Lack

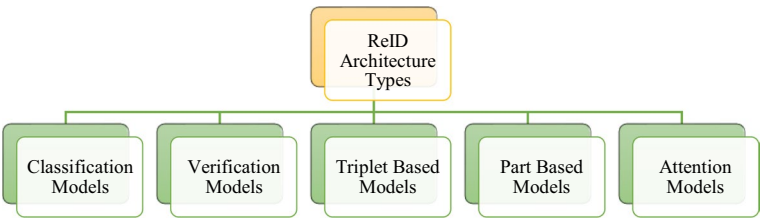


Fig. 6 Different Kinds of Deep Architectures used for Re-Id methods

of diverse samples often lead to over-fitting. The softmax loss is usually employed in classification models which encourages the separation of different identity samples [62, 172]. However, Re-Id presents large intra-class variations such as pose variations, view variations, lightning variations, occlusion, background clutter etc., for which the softmax loss has performed poorly. Several improvements have been suggested over the softmax loss to handle these intra-class challenges.

Zhu et al. [172] aim to overcome the inability of softmax loss to handle intra-class variations by using it in conjunction with the center loss [121] which was originally used for facial recognition. Authors train a convolutional neural network (CNN) with the proposed combination of softmax and center loss to extract discriminative features and obtain better Re-Id results.

Zhong et al. [168] enhance the Re-Id classification performance by using a multi-loss training setup having a combination of softmax loss, center loss and a novel "inter-center loss". While the softmax loss differentiates between different identity samples, the center loss brings same class identities closer to their center and the inter-center loss maximizes the distance between centers of different identity.

Fan et al. [24] propose a novel "Sphere Softmax Loss" by modifying the softmax loss. Instead of mapping sample images to a Euclidean space embedding, sphere loss maps sample images to the surface of a hypersphere manifold, thereby limiting spatial distribution of data points to angular variations. The proposed loss trains using the angle between sample vector and target class vector.

4.1.2 Verification models for Re-Id

Verification Models consider Re-Id to be a binary-classification problem. Given a pair of images, these models classify them to be either same or different. These models implement a pair of CNNs to extract features from input pair and compare their similarity [157]. Verification models use the contrastive loss [36] which was first used for dimensionality reduction. In Re-Id, the contrastive loss tries to pull same identity pairs to zero distance in the feature space while pushing different identity pairs beyond a given margin. Verification models suffer from the *class imbalance problem*. Consider a dataset having K number of identities, each of which has M number of image samples. The dataset is balanced with respect to different identity samples. However, if we consider the number of positive and negative samples present with respect to a single identity, there are M positive samples and (K-1)M number of negative samples. This leads to class imbalance in training verification based models.

Zhang et al. [159] combine the verification and classification Re-Id models to learn "deep features from body and parts" (DFBP). Specifically, the verification model is implemented using two neural networks that are trained by comparing body parts from image sub-regions of input pair while the global region features are extracted from a classification model to learn body-based features. The concatenated body-based and part-based features form the final representation. Zhong et al. [167] propose a novel "Feature Aggregation Network" (FAN) which also combines the classification and verification tasks. FAN extracts multi-level CNN features from input image pairs. Then, Recurrent Comparative Network (RCN) containing attention module compares appearance of input image pairs for verification loss. The CNN features are pooled directly using the Global Average Pooling (GAP) for classification loss. Figure 7 shows the proposed model.

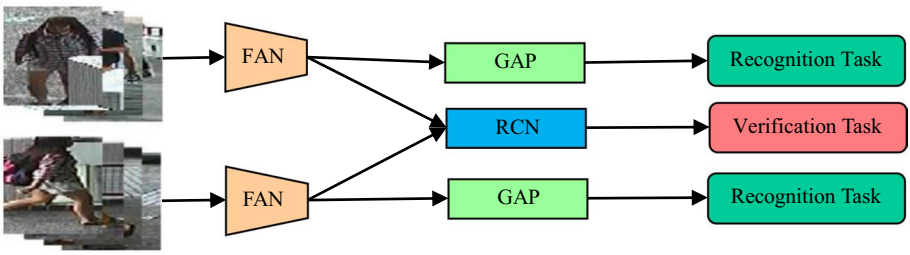


Fig. 7 Overview of proposed hybrid model. CNN features are extracted from input image pair using FAN networks. RCN learns joint representation for verification task while GAP is used for recognition task [167]

4.1.3 Triplet based Re-Id models

Triplet models for Re-Id take triplet input units. Each triplet unit contains three image samples: the anchor, a positive sample (having same identity as the anchor) and a negative sample (different identity from the anchor) [8]. The triplet loss [20] is trained to keep the Euclidean distance between anchor and positive sample less than anchor and negative sample. Let T_i represent the ith triplet such that $T_i = \langle T_i^a, T_i^p, T_i^n \rangle$ having anchor image T_i^a, positive sample T_i^p and negative sample T_i^n. $\Gamma(I)$ represents the extracted CNN features for image I. $\|x\|$ represents the $\mathcal{L}_2$ norm. The proposed triplet loss enforces the following condition:

$$\|\Gamma(T_i^a) - \Gamma(T_i^p)\| < \|\Gamma(T_i^a) - \Gamma(T_i^n)\| \tag{1}$$

Figure 8 demonstrates the preservation of condition (1).

The main drawback of triplet models is that they only use weak annotations from a triplet to learn discriminative features, unlike a classification model learning from numerous samples of a given identity available in the dataset [128]. The traditional triplet loss has

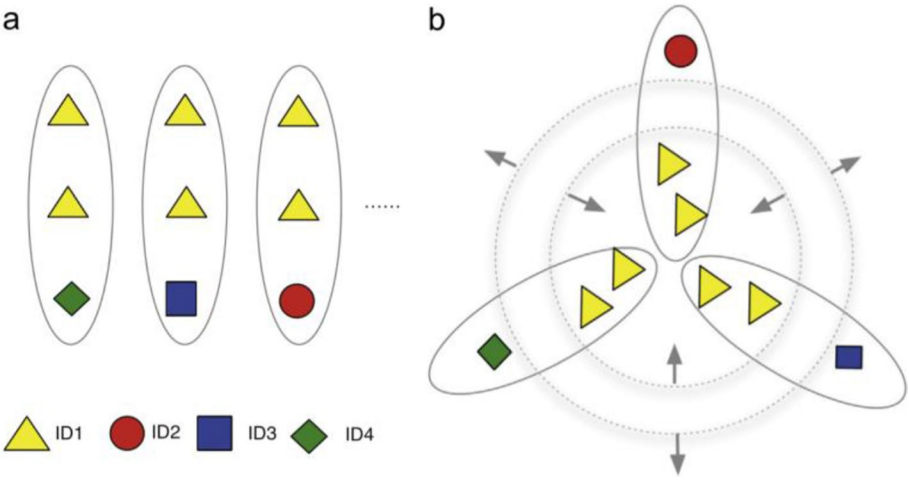


Fig. 8 **A**) Three triplet units having anchor identity (yellow triangle), positive sample (also yellow triangle) and negative sample (green/purple/red shapes). **b**) The Triplet Loss enforces the features of positive samples closer meanwhile pushing the negative samples away [20]

faced the issue of slow convergence and hence, several improvements have been formulated to improve the discrimination ability of triplet based models. Table 3 demonstrates these novel improved triplet losses.

4.1.4 Part-based Re-Id models

In the initial years, deep Re-Id methods focussed mostly on extracting global image-level features to identify individuals [35, 74, 120]. However, this approach quickly became ineffective in handling small inter-class variations such as identifying different people wearing same color clothes. This has led to a gain in popularity of part-based Re-Id methods due to their superior discrimination capability based on finer part-level cues which are usually suppressed while extracting global features [143]. Part-Based Re-Id methods extract different image regions to find discriminative part-level features.

Yan et al. [142] propose a feature attention block for part-based Re-Id. The authors slice features maps into spatial features and assign them weights thereby highlighting the important part regions as demonstrated in Fig. 9. Tian et al. [103] propose a novel Strong Part Based Network (SP_Net) that divides feature maps into $\mathbb{N}$ parts, thereby learning part level features using $\mathbb{N}$ part losses combined to obtain a local loss. The local loss is combined with global loss in a weighted manner to obtain discriminative capabilities. The major challenges faced by part-based models are the variations in pose, alignment and scale of corresponding image regions (parts) under comparison.

4.1.5 Attention-based Re-Id models

Attention based Re-Id models aim to selectively choose regions of high interest from input information [107, 120]. The proposed "Attention Modules" focus on extracting regions containing highly discriminative features while ignoring other regions having little or no discriminative capability. Such an approach of targeting specific regions helps to overcome Re-Id challenges like background clutter, misalignment etc. [69]. Attention models have proven their superior performance in several computer vision applications including Re-Id

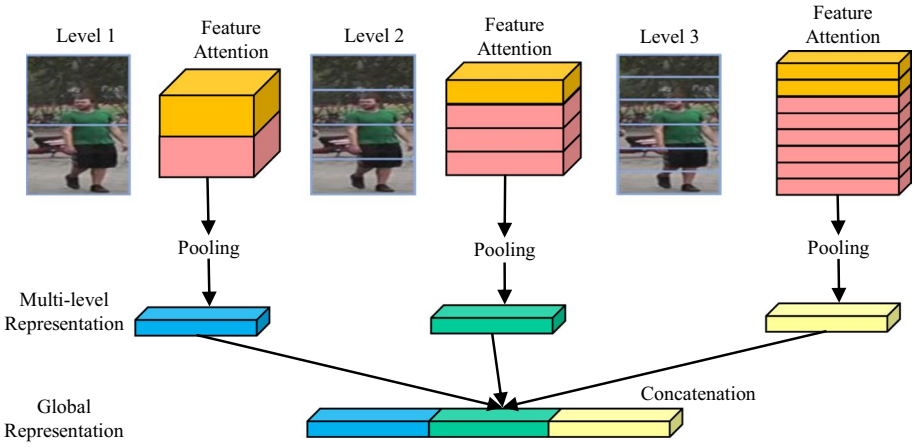


Fig. 9 Multiple feature attention networks used to generate different levels of part features. Extracted feature vectors are concatenated to obtain global representations [142]

with the growth of Recurrent Neural Networks (RNN) based on Long Short Term Memory (LSTM) [41]. Various Re-Id implementations have incorporated the attention mechanism to enhance their performance. Table 4 presents an analysis of contributions of attention modules in Re-Id methods.

Table 5 gives a detailed overview of various deep Re-Id contributions based on their architecture types.

The **highlights of deep Re-Id architectures** are:

- Deep learning based Re-Id architectures can be classification models, verification models, triplet-based models, part-based models and attention based models.
- Classification models treat Re-Id as a multi-class classification problem that use the softmax loss to predict the class of an input query. Softmax loss encourages the separation of different classes but struggles with large intra-class variations. Several Re-Id works combine classification with other model types to overcome this limitation [68, 91, 121].
- Verification models treat Re-Id as a binary classification problem, taking a pair of inputs and categorizing them as same or different. These models suffer from the class imbalance problem since the number of positive pair combinations is far less than the negative pairs where each identity contains same number of samples in the dataset.
- Triplet models input triplets of images containing an anchor, a positive and a negative sample. These models are trained on the triplet loss that aims to pull the positive sample closer and pushes the negative sample away in feature space. Several improvements have been suggested over the traditional triplet loss having convergence issues. These improvements include solving slow convergence [144], removing triplet mining overhead [39], maintaining intra-class compactness [16] etc.
- Part-based Re-Id models aim to focus on sub-regions within input to extract finer feature representations crucial in differentiating samples with small inter-class variations, usually missed in global image level representations. Focus on different parts of feature maps [103, 142], face and body regions [53], attribute guided body parts [55] etc., have yielded discriminative feature representations.
- Attention models highlight regions of high interest within input holding highly discriminative information. Gaining popularity since the growth on RNN and LSTM, attention modules have helped in implementing *spatial attention* within image/frame regions such as human body parts [6], foreground [170], physical attributes. [102] and *channel attention* within deep features [54].

4.2 Methods based on Re-Id challenges

The task of person re-identification has faced several challenges like sample variations in view, pose, lightning and scale, partial or complete occlusion [13, 140, 156], background clutter etc. Figure 10 demonstrates some of these Re-Id challenges presented by samples from various Re-Id datasets as detailed in Table 2. Any re-id system striving to achieve competent recognition rates must be able to counter these challenges effectively. Numerous research works have been conducted with this motivation.

Feng et al. [26] attempt to overcome the challenge of large intra-class disparity caused by view variations from images captured by cameras placed at different viewpoints. Authors propose a framework capable of learning view-specific features consistent with

Fig. 10 Re-Id challenges posed by dataset samples. Each column demonstrates samples from a unique identity. Columns (i) and (ii) show variations in pose, (iii) and (iv) show scale variations, (v) and (vi) display occlusion examples, (vii) and (viii) describe background clutter, (ix) and (x) show view variations in samples. All samples have been derived from the Market-1501 dataset

each camera view which utilizes a cross-view Euclidean constraint (CV-EC) and cross-view center loss to decrease the distance of features of same person from different views. Qi et al. [85] tackle lightning and viewpoint variation at two levels. Firstly, they extract deep features trained on multiple datasets which are robust to differences in illumination and view. Secondly, they use the learned features to find an optimal ensemble of metrics including the Cosine distance metric that reduces the intra-class disparity even further. Sikdar et al. [92] achieve scale-invariance by modifying the convolution functionality within a deep network. Instead of learning a kernel on a fixed scale input, the input is first transformed to a pyramid of multiple resolutions. The network then learns multiple scaled feature maps which are then re-scaled to original size before applying max pooling. Such an operation has proven to produce scale invariant results for re-id system. Input image misalignment can seriously hamper the feature learning and matching process. To handle the misalignment problem, Zheng et al. [163] introduce *pose invariant embedding* (PIE) which aligns identities within sample images to a standard pose using pose estimation. The transformed standard pose promotes learning of discriminative feature extraction and matching and is alignment invariant. Table 6 presents some novel Re-Id contributions that are robust to Re-Id challenges.

The **highlights of contributions based on Re-Id challenges** are:

- Finding a person of interest is challenging due to visual variations in pose, view, lightning, scale, partial or complete occlusion, background clutter and misalignment.
- Several deep Re-Id contributions have aimed to develop robust methodologies against these Re-Id challenges.
- Skeleton joints data and clothe colors have produced pose and lightning invariance [146], learning view-specific representations for view invariance [14], utilizing foreground attentive network to suppress noisy background [170], convolving with multi-scale input to obtain scale invariant features [92] and using pose estimation to achieve pose invariance [163] are some of efforts to overcome these challenges.

4.3 Re-Id methods based on modality

Visible images have proven to be the most common source of discriminative information crucial for identifying individuals. Hence, research literature is filled with visible image based Re-Id methods due to their superior identification power. Despite their popularity, the visible image based methods are prone to several challenges already discussed in Sect. 4.2. Hence, some cross-modal Re-Id methods have also been proposed to enhance the ability of Re-Id systems further. This section discusses various visible light based and cross-modality based Re-Id methods.

4.3.1 Visible image based Re-Id methods

Numerous visible image based Re-Id contributions have focussed upon extracting discriminative deep features to achieve high recognition rates for re-identification [37, 105, 112, 115]. Table 7 details some novel Re-Id image based contributions.

Wu et al. [130] propose a five-branch deep model that is capable of learning features not only from the usual horizontal direction but also in the vertical direction. The model scans for spatial information of body parts from left to right and head to foot, thereby learning discriminative information. Working with one type of features is often limiting in finding discriminative capabilities for re-id systems. Hence, an obvious direction of improvement is to fuse different features together to obtain higher differentiation in re-id. Zhao et al. [161] introduce a novel deep triplet model (MT-net) performing multi-level feature extraction. Both detailed and global features from each layer are combined together in an optimal proportion through training. The fused features prove to have high discriminative capabilities. While most deep learning based re-id approaches extract features only from the top layer, middle layer features can also contribute to discriminative capabilities of the model in certain situations. Wu et al. [131] introduce a *multi-level feature network with multiple losses* (MFML) which is a multi-branch architecture representing multiple middle layer representations trained on triplet loss and top layer representation trained on the hybrid loss. The representation from various layers are fused in a weighted manner based on their importance in obtaining differentiating characteristics. Based on the idea that color features hold key information in reference to re-id task, Liu et al. [73] fuse traditional Gaussian of Gaussian (GOG) features from four color channels (RGB, Lab, HSV, RnG) with deep features to obtain highly discriminative features which achieve state-of-the-art performances.

4.3.2 Cross-modality Re-Id methods

While image based methods have proven to be most popular in the Re-Id research community, these visual feature based methods have several limitations. Visual Re-Id methods face several challenges such as variations in pose, view, lightning, scale, partial or complete occlusion, background clutter which are already discussed in Sect. 4.2. Also, visual Re-Id methods are highly ineffective in dark environments such as nighttime, due to the poor illumination that suppresses most of the visual cues [27]. This has led to the development of multi-modal Re-Id methods [72, 72] that combine data from multiple modalities to reduce dependency on visual information. Table 8 lists the

novel multi-modal Re-Id methods combining RGB, depth, text and infrared (thermal) based data modalities.

Chang et al. [9] learn a similarity metric for visual and textual representations. Authors use the Resnet architecture to extract visual features from image samples. A 2500-dimensional textual feature embedding is extracted from textual description of each sample (gender, clothe color etc.) using tokenization, lemmatization and stemming. The model is trained in an end to end manner using a triplet architecture on both visual and textual representations.

Gait information has proven to be highly discriminative for Re-Id tasks. Gohar et al. [32] formulate a multi-modal Re-Id method performing non-visual gait analysis on information extracted using wearable sensors namely, accelerometer and gyroscope. The proposed method learns discriminative information integrating the temporal aspect of gait data using Recurrent Gated Units (GRU).

Other significant multi-modal approach is [59].

The **highlights of modality based deep Re-Id contributions** are:

- Visible RGB images are the most widely used data modality for deep Re-Id methods due to the rich variety of visual information they contain and the growth of several image based Re-Id datasets developed over the years.
- Visible RGB image methods aim to extract discriminative features by using multi-branch networks [130], fusing multiple features together [161], proposing novel loss functions [113] etc.
- Visible RGB image methods face several visual challenges as discussed in Sect. 4.2 and lose most of their discriminative capabilities in dark/night environments.
- Multi-modal Re-Id methods reduce the dependency on visual information for extracting discriminative features.
- Several multi-modal Re-Id contributions have extracted anthropometric features from depth data [89], modality specific representations from thermal images [137], gait information from accelerometer and gyroscope [32] and combined with visual information to formulate multi-modal representations.

4.4 Cross-domain Re-Id methods

Based on the kinds of samples present, different re-id datasets hold different generalizations of human appearance. Hence a re-id model trained on one dataset performs poorly on a different dataset. Several works have addressed this issue with domain adaptation techniques [19, 75, 95] attempting to bridge the gap between the learned source domain to the unknown target domain as demonstrated in Table 9. Zhang et al. [160] introduce a novel *Dual Generation Learning* (DGL) method for unsupervised domain adaption such that a re-id model when evaluated on any relevant dataset shows acceptable recognition results. The DGL method generates target style images for samples from source dataset and camera style images for those from target dataset, thereby expanding them to consider varying domain styles. Ganin et al. [29] propose to augment few standard layers and a novel *gradient reversal layer* into deep architectures to learn features that are trained on both labelled source domain as well as unlabelled target domain. Such features cannot discriminate between source and target domain and hence are suitable for domain adaptation. Wang et al. [109] propose a *Deep Multi-Task Transfer Network* (DMTNet) to transfer discriminative features learnt from source domain to target domain by utilizing a cluster estimating

algorithm, attribute attention important learning and multi-task learning. Other domain adaptation based re-id works involve refining learned augmented attribute features according to target domain [141], image to image translation using generative adversarial network [171], viewpoint transfer using generative adversarial network [99] etc.

The **highlights of cross-domain Re-Id methods** are:

- Most deep Re-Id methods that are trained/tested on few datasets perform poorly on other datasets.
- Several cross-domain Re-Id approaches have been developed to either learn better generalization across multiple datasets or transfer learned characteristics of source domain to another target domain.
- The cross-domain approaches have achieved sample style transfer from source to target dataset [160], training from multiple dataset combinations [30], transformation of learned features from labelled source to unlabelled target dataset [141], thereby improving the generalizability of deep Re-Id methods.

4.5 Metric learning methods for Re-Id

Metric learning has proven to be a significant step in computer vision problems such as person re-identification, face recognition etc. [175]. Metric learning aims to find a similarity function on extracted features that is used to decrease positive pair distance while increasing negative pair distance [5]. Table 10 highlights the different kinds of metric learning contributions reviewed in this article. Since the underlying data distributions varies with the nature of the computer vision task, the ideal similarity function is mostly task-specific [43]. Traditional metric learning methods focussed on learning linear Mahalanobis based metrics. Such linear metrics utilized the sample distribution centroid and standard deviation to evaluate the similarity with a give sample point. However, such linear metrics often failed to comprehend nonlinear relationships among the samples.

Recently, deep metric learning techniques have become increasingly popular due to their ability to capture hierarchal nonlinear affiliations. Hu et al. [43] propose a *deep transfer metric learning* (DTML) method that transfers discriminative information from labelled source domain to unlabelled target domain. DTML learns hierarchical nonlinear transformations, maximizing the inter-class disparity, minimizing the intra-class differences and restricting the divergence of source and target domain. Ding et al. [21] increase the generalizability of metric learning by introducing *robust discriminative metric learning* (RDML) which is insensitive to noise unlike most metric learning techniques. A fast low rank model is also used to discover global structure within data and ensure scalability to larger datasets. Lin et al. [67] argue that similarity transformations cannot capture deformations like translation and shearing in cross domain visual matching tasks. The authors propose *a generic similarity metric* that generalizes the similarity transformation to affine transformation capable of capturing complex deformations. While most metric learning techniques focus on the overall similarity learning of samples, Duan et al. [22] propose the *deep localised metric learning* (DLML) to learn multiple metrics fine-grained over numerous local subspaces. DLML specializes in handling data varying locally. Hu et al. [44] introduce the *sharable and individual multi-view deep metric learning* (MvDML) to utilize multi-view data in best possible manner. MvDML learns the best combination of distance metrics by focussing on both individual view specifics metric as well as a combined

multi-view representation. Chen et al. [15] propose a novel *pose invariant deep metric learning* (PIDML) method that utilizes pose invariant embedding [163] and an improved triplet loss to achieve pose invariance for metric learning. Ren et al. [88] propose a *deep structured metric learning method* that utilizes a novel structured loss function to achieve robust person re-identification. The proposed loss skips positive sample pairs of small distance and negative sample pairs having large distance.

The **highlights of deep metric learning based Re-Id methods** are:

- Metric learning means to learn a similarity function that pulls features from same identity samples closer and pushes different identity features away.
- Deep learning based metric learning have gained popularity in recent years due to their ability to learn nonlinear sample associations unlike the traditional Mahalanobis metric learning only linear relationships.
- Several novel deep metric learning approaches have been proposed that give better discrimination capability than the usual Euclidean or Cosine measure [174], can perform cross-domain learning [43], and are robust to noise [21], deformation [67] etc.

5 Video based deep Re-Id contributions

This section explores the video based deep person re-identification methods.

Video Re-Id methods received less attention compared to image based techniques in the initial years of Re-Id research [129]. A major reason was the lack of large-scale video datasets. However, recent years have witnessed the growth of several video based Re-Id datasets such as [40, 83, 118, 162] encouraging the development of video Re-Id methods. Compared to images, video sequences have advantageous characteristics that can be exploited for Re-Id tasks. *Firstly*, a video sequence comprising of multiple frames holds essential temporal information across frames giving crucial motion cues which are absent from image samples. *Secondly*, multiple frames from each video provide numerous visual examples per identity adding diversity in samples. *Thirdly*, while image based methods mostly focus on exploiting visual features, video Re-Id methods target a combination of visual and temporal features to extract discriminative features for Re-Id tasks which prove to be more robust to Re-Id challenges. While video based methods give rise to more discriminative feature embeddings, they also add to the existing Re-Id challenges as demonstrated in Fig. 11. *Firstly*, since the simplest way of generating video features is by fusing frame level features together and frame level features are based on 2D convolution operations that totally neglect the temporal aspect of frame sequences, the fusion operation misses all of the temporal cues crucial for video Re-Id. *Secondly*, videos have different frame rates and

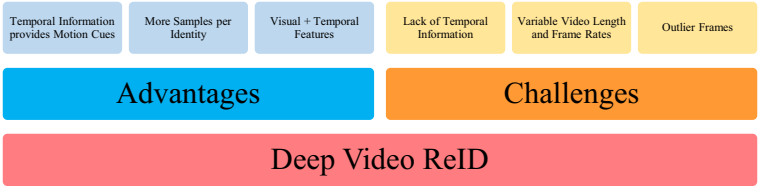


Fig. 11 Advantages and Challenges of Video based Person Re-Identification methods

different time series that makes it hard to make comparisons among samples. *Thirdly*, not all frames provide discriminative information and some outlier frames prove misleading in learning robust video representations.

Recent developments in Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models [41] have boosted video Re-Id by providing the capability of extracting motion cues from temporal information for robust video representations. McLaughlin et al. [79] propose a video re-identification system based on recurrent neural network for wide area tracking. The proposed network uses a recurrent layer to combine frame-level details from all frames into a single combined appearance feature representing the entire video showing competent recognition results. Attention modules have played key role in video Re-Id to identify and isolate outlier frames. They also help in extracting discriminative regions within video frames. Wu et al. [129] propose a Siamese attention network that learns to realize which regions (where) from which frames (when) are relevant for comparing identities. The attention mechanism learns the most relevant features by focussing on distinct regions helping to identify given identity. Zhang et al. [152] propose a *self and collaborative attention network* (SCAN) for video re-id. The proposed model takes a pair of videos as input, aligns and compares their discriminative frames using a generalized similarity measurement module and refines intra-sequence and inter-sequence features from videos using a non-parametric attention module. Wu et al. [132] argue that merely combining frame-wise features to obtain overall video features is ineffective as the temporal cues get lost in 2D convolutions. Authors propose a novel 3D Person VLAD Aggregation layer that helps to extract appearance and motion characteristics of entire video. The model also handles occlusions and misalignments through soft attention modules. Zhang et al. [154] propose a *multi-scale spatial–temporal attention network* (MSTA) that focusses attention to regions within video frames at different scales. It contains a ResNet50 based encoder responsible for extracting frame-level features from discriminative regions and an aggregator to fuse features from different scales Fig. 12.

Table 11 dives into several contributions highlighting the diverse approaches towards video person re-identification using deep learning.

6 Conclusion and future directions

6.1 Conclusion

This review provides a comprehensive and exhaustive analysis of deep learning based person re-identification methods. The objective of this work is to give the readers a thorough understanding of different approaches towards deep Re-Id. This reviewed literature has been divided into several logical categories as demonstrated in the taxonomy diagram Fig. 2. These approaches have been classified on the basis of adopted architecture types (classification, verification, triplet based, part based and attention models) integrating

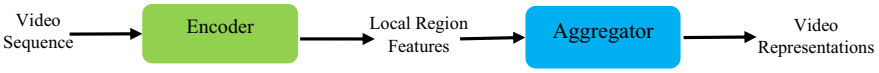


Fig. 12 Architecture of proposed MSTA model. The encoder extracts framewise features and the aggregator fuses the frame-level features to obtain video representations [154]

different kinds of losses (softmax, triplet), the common Re-Id challenges faced (variations in pose, lightning, view, scale, partial or complete occlusion, background clutter), image based methods and multi-modal Re-Id methods reducing the dependency on visible RGB approaches, cross-domain methods to improve generalizability of approaches across different datasets, metric learning approaches to learn ideal similarity functions and deep video Re-Id methods exploiting both spatial and temporal cues from multiple frames of a video sequence. Each category presented as a separate section provides an extensive look into these contribution types and the highlights at the end of each section gives a quick overview of the reviewed methods. Tables 3, 4, 5, 6, 7, 8, 9, 10, 11 provide the *key ideas* behind numerous deep Re-Id methods across various categories.

Part-Based and Attention architectures are the more popular architecture types in recent times owning to their ability to find regions of rich information and extraction of finer visual cues. Re-Id challenges have been conquered using different techniques like using pose estimation to achieve pose invariance [163], obtain scale invariance through multi-scale input convolutions [92] etc. Some multi-modal approaches have helped to reduce the over dependence on visible RGB based methods making them more viable for darker environments with fewer visual cues [137]. Deep video Re-Id methods provide more discriminative information when compared to image based approaches due the motion cues present in temporal information across frames. However, video Re-Id presents its own challenges of finding discriminative regions both spatially (within frames) and temporally (across frames) as demonstrated in Fig. 11.

6.2 Future directions

This section presents the possible future research directions for the benefit of the readers in deciding upcoming research possibilities.

Transformer-based Approaches: The recent rise in the popularity of transformer achitectures proves their effectiveness in solving computer vision problems. Novel transformer models can be used to extract more discriminative features leading to better identification results6.
Multi-Modal Information Fusion: Fusing discriminative information from multiple modalities such as visual, depth, thermal images etc. can be a great way of improving the discrimination capabilities of deep person re-identification networks. There is also a need to develop multi-modal reid datasets that help to build such approaches 7.
Cross-Modality Query Search: Real-world person re-identification systems requires cross modality query search facility. This means that a query can be an image while it is being matched with a video dataset. Reid approaches should be capable of searching the same person in a different modality8.
Cross-Domain Adaptation: The performance of person re-identification systems degrade when they are trained on one dataset and tested on another. This is due to the changing degree of variations in lightning, pose, alignment, scale etc., in different person re-identification datasets. More focus can be given in developing robust and generalized cross-domain approaches of person re-identification that perform well across multiple datasets9.
Privacy Preservation: Person re-identification has widespread applications for surveillance in public places. Hence, maintaining the anonymity of individuals is a key issue. Future research works can focus on developing techniques that preserve the identity of individuals in public spaces while deploying person re-identification systems.
Real-World Deployment: Person re-identification approaches suffer in real world settings. Hence, there is an urgent need to build robust person re-identification systems that

work well in realistic scenes like [49] and aid in the utilization of person re-identification in retail, security and law enforcement.

Long-Term Appearance Variation: While most reid approaches have focussed on common reid challenges such as variations in pose, lightning etc., very few methods have focussed on identifying the long-term variations of a person's appearance and clothing style. This includes variations of facial attributes, dressing etc. Robust reid systems are needed that can identify the same person after a long period of time.

Semi-Supervised and Unsupervised Methods: While supervised methods have scored better than semi or unsupervised reid methods, creating labelled datasets for supervised approaches is a huge research challenge. Hence, more novel semi and unsupervised reid approaches are needed that bridge the performance gap between them and supervised methods.

Federated Learning based ReID: Since reid has critical applications in surveillance based tasks, federated learning is a promising upcoming paradigm that allows for decentralized training of models from non-overlapping geographical regions for better recognition capabilities.

Other Possible Improvements: There are few end-to-end Re-Id research contributions that involve both person detection and re-identification together in a single framework. Since most datasets are collected under controlled environment, the person detection is usually performed automatically. Re-Id can also be used for continuous authentication purposes [61, 63, 64, 65]. End-to-end Re-Id is a promising research direction. GAN based Re-Id works have greatly supported the style transfer requirements of cross-domain Re-Id approaches. However, the low to medium quality of generated samples has limited the performance of these approaches. An improvement in sample generation quality can significantly boost cross-domain Re-Id approaches. Most part-based Re-Id contributions have focussed on systematic comparison between corresponding part regions of input pairs. But the contextual relationship among different regions is mostly ignored. Preserving the semantic relationship among different parts like [3] is a potential way of improving part-based methods further. Attribute-based methods have become increasingly popular in finding finer visual cues. These attribute-based methods can be extended further for applications like part localization like in [55] to achieve better Re-Id performance.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Almasawa MO, Elrefaei LA, Moria K (2019) A Survey on Deep Learning-Based Person Re-Identification Systems. IEEE Access 7:175228–175247
2. Avola D, Cinque L, Fagioli A, Foresti GL, Pannone D, Piciarelli C (2020) Bodyprint—A Meta-Feature Based LSTM Hashing Model for Person Re-Identification. Sensors 20(18)
3. Bai X, Yang M, Huang T, Dou Z, Yu R, Xu Y (2020) Deep-Person: Learning discriminative deep features for person Re-Identification. Pattern Recogn 98
4. Baltieri D, Vezzani R, Cucchiara R (2011) 3DPeS: 3D people dataset for surveillance and forensics. In: Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding, Scottsdale, Arizona, USA.

5. Bąk S, Carr P (2018) Deep deformable patch metric learning for person re-identification. IEEE Trans Circuits Syst Video Technol 28(10):2690–2702
6. Bao T, Wang B, Karmoshi S, Liu C, Zhu M (2019) Learning Discriminative Features through an Individual's Entire Body and The Visual Attentional Parts for Person Re-Identification. Int J Innov Comput Inf Control 15(3):1037–1048
7. Barbosa IB, Cristani M, Bue AD, Bazzani L, Murino V (2012) Re-identification with RGB-D Sensors. European Conference on Computer Vision (ECCV) Workshop, Florence, Italy
8. Cai X, Liu L, Zhu L, Zhang H (2021) Dual-modality hard mining triplet-center loss for visible infrared person re-identification. Knowl-Based Syst 215
9. Chang Y-S, Wang M-Y, He L, Lu W, Su H, Gao N, Yang X-A (2020) Joint deep semantic embedding and metric learning for person re-identification. Pattern Recogn Lett 130:306–311
10. Chen G, Lin C, Ren L, Lu J, Zhou J (2019) Self-critical attention learning for person re-identification. In: International Conference on Computer Vision (ICCV), Seoul
11. Chen S-Z, Guo C-C, Lai J-H (2016) Deep Ranking for Person Re-Identification via Joint Representation Learning. IEEE Trans Image Process 25(5):2353–2367
12. Chen Y, Duffner S, Stoian A, Dufour J-Y, Baskurt A (2018) Deep and low-level feature based attribute learning for person re-identification. Image Vis Comput 79:25–34
13. Chen Y, Yang Y, Liu W, Huang Y, Li J (2022) Pose-guided counterfactual inference for occluded person re-identification. Image Vis Comput 128
14. Chen Y-C, Zhu X, Zheng W-S, Lai J-H (2018) Person Re-Identification by Camera Correlation Aware Feature Augmentation. IEEE Trans Pattern Anal Mach Intell 40(2):392–408
15. Chen M, Ge Y, Feng X, Xu C, Yang D (2018) Person Re-Identification by Pose Invariant Deep Metric Learning With Improved Triplet Loss. IEEE Access 6:68089–68095
16. Cheng D, Gong Y, Zhou S, Wang J, Zheng N (2016) Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas
17. Cheng DS, Cristani M, Stoppa M, Bazzani L, Murino V (2011) Custom Pictorial Structures for Re-identification. In: British Machine Vision Conference, Dundee, UK
18. Choe C, Choe G, Wang T, Han S, Yuan C (2019) Deep feature learning with mixed distance maximization for person re-identification. Multim Tools Applic 78:27719–27741
19. Delussu R, Putzu L, Fumera G (2023) Human-in-the-loop cross-domain person re-identification. Expert Syst Applic 226
20. Ding S, Lin L, Wang G, Chao H (2015) Deep feature learning with relative distance comparison for person re-identification. Pattern Recogn 48(10):2993–3003
21. Ding Z, Shao M, Hwang W, Suh S, Han J-J, Choi C, Fu Y (2019) Robust Discriminative Metric Learning for Image Representation. IEEE Trans Circuits Syst Video Technol 29(11):3173–3183
22. Duan Y, Lu J, Feng J, Zhou J (2018) Deep localized metric learning. IEEE Trans Circ Syst Video Technol 28(10):2644–2656
23. Ess A, Leibe B, Gool LV (2007) Depth and appearance for mobile scene analysis. In: In International Conference on Computer Vision (ICCV), Rio de Janeiro
24. Fan X, Jiang W, Luo H, Fei M (2019) SphereReID: Deep hypersphere manifold embedding for person re-identification. J Vis Commun Image Represent 60:51–58
25. Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D (2010) Object Detection with Discriminatively Trained Part-Based Models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
26. Feng Z, Lai J, Xie X (2018) Learning view-specific deep networks for person re-identification. IEEE Trans Image Process 27(7):3472–3483
27. Feng Z, Lai J, Xie X (2019) Learning modality-specific representations for visible-infrared person re-identification. IEEE Trans Image Process 29:579–590
28. Fu M, Sun S, Chen N, Wang D, Tong X (2019) Deep Fusion Feature Presentations for Nonaligned Person Re-Identification. IEEE Access 7:73253–73261
29. Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, Marchand M, Lempitsky VS (2016) Domain-adversarial training of neural networks. J Mach Learn Res 17(1)
30. Genç A, Ekenel HK (2019) Cross-dataset person re-identification using deep convolutional neural networks: effects of context and domain adaptation. Multim Tools Applic 78:5843–5861
31. Girija S, Baker T, Ahmed N, Khedr AM, Aghbari ZA, Jha A, Sobolev K, Asl SA, Phan A-H (2023) Attribute recognition for person re-identification using federated learning at all-in-edge. Internet Things 2023
32. Gohar I, Riaz Q, Shahzad M, Ul MZ, Hashmi H, Tahir H, Ul Haq ME (2020) Person re-identification using deep modeling of temporally correlated inertial motion patterns. sensors 20(3)

33. Gou M, Karanam S, Liu W, Camps O, Radke RJ (2017) DukeMTMC4ReID: A Large-Scale Multi-camera Person Re-identification Dataset. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA.
34. Gray D, Brennan S, Tao H (2007) Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International Workshop on Performance Evaluation for Tracking and Surveillance, Rio de Janeiro
35. Gupta A, Pawade P, Balakrishnan R (2022) Deep Residual Network and Transfer Learning-based Person Re-Identification. Intell Syst Applic 16
36. Hadsell R, Chopra S, LeCun Y (2006) Dimensionality reduction by learning an invariant mapping. In: IEEE Conference on Computer Vision and Pattern Recognition, New York
37. He L, Du H, Fu Y, Ye Y (2023) Multiple camera styles learning for unsupervised person re-identification. Optik 277
38. He Z, Jung C, Fu Q, Zhang Z (2019) Deep feature embedding learning for person re-identification based on lifted structured loss. Multim Tools Applic 78:5863–5880
39. Hermans A, Beyer L, Leibe B (2017) In Defense of the Triplet Loss for Person Re-Identification. https://arxiv.org/abs/1703.07737
40. Hirzer M, Beleznai C, Roth PM, Bischof H (2011) Person re-identification by descriptive and discriminative classification. In: Scandinavian conference on image analysis, Ystad
41. Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. Neural Comput 9(8)
42. Hou R, Ma B, Chang H, Gu X, Shan S, Chen X (2019) VRSTC: occlusion-free video person re-identification. In: Conference on computer vision and pattern recognition, California
43. Hu J, Lu J, Tan Y-P, Zhou J (2016) Deep Transfer Metric Learning. IEEE Trans Image Process 25(12):5576–5588
44. Hu J, Lu J, Tan Y-P (2018) Sharable and Individual Multi-View Metric Learning. IEEE Trans Pattern Anal Mach Intell 40(9):2281–2288
45. Huang Y, Sheng H, Zheng Y, Xiong Z (2017) DeepDiff: Learning deep difference features on human body parts for person re-identification. Neurocomputing 241:191–203
46. Islam K (2020) Person search: new paradigm of person re-identification: A survey and outlook of recent works. Image and Vision Computing 101
47. Jiang M, Leng B, Song G, Meng Z (2020) Weighted triple-sequence loss for video-based person re-identification. Neurocomputing 381:314–321
48. Junior JCJ, Baró X, Escalera S (2018) Exploiting feature representations through similarity learning, post-ranking and ranking aggregation for person re-identification☆. Image Vis Comput 79:76–85
49. Karanam S, Gou M, Wu Z, Borras AR, Camps O, Radke RJ (2019) A Systematic evaluation and benchmark for person re-identification: features, metrics, and datasets. IEEE Trans Pattern Anal Mach Intell 41(3):523–536
50. Ke Q, Bennamoun M, Rahmani H, An S, Sohel F, Boussaid F (2018) Identity Adaptation for Person Re-Identification. IEEE Access 6:48147–48155
51. Khatun S, Denman SS, Fookes C (2023) Pose-driven attention-guided image generation for person re-Identification. Pattern Recognition 137
52. Ksibi S, Mejdoub M, Amar CB (2019) Deep salient-Gaussian Fisher vector encoding of the spatio-temporal trajectory structures for person re-identification. Multim Tools Applic 78:1583–1611
53. Koo JH, Cho SW, Baek NR, Kim MC, Park KR (2018) CNN-based multimodal human recognition in surveillance environments. Sensors 18(9)
54. Li R, Zhang B, Kang D-J, Teng Z (2019) Deep attention network for person re-identification with multi-loss. Comput Electr Eng 79
55. Li S, Yu H, Hu R (2020) Attributes-aided part detection and refinement for person re-identification. Pattern Recognition 97(107016)
56. Li W, Wang X (2013) Locally Aligned Feature Transforms across Views. In: IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA
57. Li W, Zhao R, Xiao T, Wang X (2014) DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA
58. Li W, Zhao R, Wang X (2012) Human reidentification with transferred metric learning. In: Asian conference on computer vision. Daejeon, Korea
59. Li Y, Chen Y (2022) Infrared-visible cross-modal person re-identification via dual-attention collaborative learning. Signal Process Image Commun 109
60. Li Y, Jiang X, Hwang J-N (2020) Effective person re-identification by self-attention model guided feature learning. Knowled-Based Syst 187
61. Li Y, Liu L, Qin H, Deng S, El-Yacoubi MA, Zhou MA (2022) Adaptive Deep Feature Fusion for Continuous Authentication with Data Augmentation. IEEE Trans Mobile Comput 1-16
62. Li Y, Liu L, Zhu L, Zhang H (2021) Person re-identification based on multi-scale feature learning. Knowl-Based Syst 228

63. Li Y, Luo J, Deng S, Zhou G (2022) CNN-based continuous authentication on smartphones with conditional Wasserstein generative adversarial network. IEEE Internet Things J 9(7):5447–5460
64. Li Y, Luo J, Deng S, Zhou G (2023) SearchAuth: Neural Architecture Search based Continuous Authentication Using Auto Augmentation Search. ACM Trans Sensor Netw
65. Li Y, Tao P, Deng S, Zhou G (2021) DeFFusion: CNN-based continuous authentication using deep feature fusion. ACM Trans Sensor Netw 18(2):1–20
66. Li Y, Zhang H, Liu L (2022) HCFN: Hierarchical cross-modal shared feature network for visible-infrared person re-identification. J Vis Commun Image Representation 89
67. Lin L, Wang G, Zuo W, Feng X, Zhang L (2017) Cross-Domain Visual Matching via Generalized Similarity Measure and Feature Learning. IEEE Trans Pattern Anal Mach Intell 39(6):1089–1102
68. Ling H, Wang Z, Li P, Shi Y, Chen J, Zou F (2019) Improving person re-identification by multi-task learning. Neurocomputing 347:109–118
69. Liu H, Feng J, Qi M, Jiang J, Yan S (2017) End-to-end comparative attention networks for person re-identification. IEEE Trans Image Process 26(7):3492–3506
70. Liu J, Liu J, Zhang Q (2023) M2FINet: modality-specific and modality-shared features interaction network for RGB-IR person re-identification. Comput Vis Image Underst
71. Liu J, Zheng-Jun Zha Z-J, Chen X, Wang Z, Zhang Y (2019) Dense 3D-Convolutional Neural Network for Person Re-Identification in Videos. ACM Trans Multim Comput Commun Applic 15(15)
72. Liu Q, He X, Teng Q, Qing L, Chen H (2023) BDNet: A BERT-based dual-path network for text-to-image cross-modal person re-identification. Pattern Recogn 141
73. Liu Y, Song N, Han Y (2019) Multi-cue fusion: Discriminative enhancing for person re-identification. J Vis Commun Image Represent 58:46–52
74. Liu Y, Zhang Y, Bhanu B, Coleman S, Kerr D (2021) Multi-level cross-view consistent feature learning for person re-identification. Neurocomputing 435:1–14
75. Liu Z, Feng C, Chen S, Hu J (2023) Knowledge-Preserving continual person re-identification using Graph Attention Network. Neural Netw 161:105–115
76. Loy CC, Xiang T, Gong S (2010) Time-delayed correlation analysis for multi-camera activity understanding. Int J Comput Vision 90:106–129
77. Lu Y, Jiang M, Liu Z, Mu X (2023) Dual-branch adaptive attention transformer for occluded person re-identification. Image Vis Comput 131
78. Luo H, Jiang W, Zhang X, Fan X, Qian J, Zhang C (2019) AlignedReID++: dynamically matching local information for person re-identification☆. Pattern Recogn 94:53–61
79. McLaughlin N, Rincon JMd, Miller P (2019) Video Person Re-Identification for Wide Area Tracking Based on Recurrent Neural Networks. IEEE Trans Circ Syst Video Technol 29(9):2613–2626
80. Meng J, Wu A, Zheng W-S (2019) Deep asymmetric video-based person re-identification. Pattern Recogn 93:430–441
81. Nguyen DT, Hong HG, Kim KW, Park KR (2017) Person recognition system based on a combination of body images from visible light and thermal cameras. Sensors 17(3)
82. Pala F, Satta R, Fumera G, Roli F (2016) Multimodal Person Reidentification Using RGB-D Cameras. IEEE Trans Circuits Syst Video Technol 26(4):788–799
83. Pan H, Chen Y, He Z (2023) Multi-granularity graph pooling for video-based person re-identification. Neural Netw 160:22–33
84. Perwaiz N, Fraz MM, Shahzad M (2018) Person re-identification using hybrid representation reinforced by metric learning. IEEE Access 6:77334–77349
85. Qi M, Han J, Jiang J, Liu H (2019) Deep feature representation and multiple metric ensembles for person re-identification in security surveillance system. Multimed Tools Appl 78:27029–27043
86. Qian X, Fu Y, Xiang T, Jiang Y-G, Xue X (2020) Leader-based multi-scale attention deep architecture for person re-identification. IEEE Trans Pattern Anal Mach Intell 42(2):371–385
87. Quispe R, Pedrini H (2019) Improved person re-identification based on saliency and semantic parsing with deep neural network models. Image Vis Comput 92
88. Ren C-X, Xu X-L, Lei Z (2019) A deep and structured metric learning method for robust person re-identification. Pattern Recognit 96
89. Ren L, Lu J, Feng J, Zhou J (2017) Multi-modal uniform deep learning for RGB-D person re-identification. Pattern Recogn 72:446–457
90. Ren L, Lu J, Feng J, Zhou J (2019) Uniform and variational deep learning for RGB-D object recognition and person re-identification. IEEE Trans Image Process 28(10):4970–4983
91. Shen C, Jin Z, Chu W, Jiang R, Chen Y, Qi G-J, Hua X-S (2019) Multi-level Similarity Perception Network for Person Re-identification. ACM Trans Multim Comput Commun Applic 15(2)
92. Sikdar A, Chowdhury AS (2020) Scale-invariant batch-adaptive residual learning for person re-identification. Pattern Recogn Lett 129:279–286

93. Si T, Zhang Z, Liu S (2019) Compact Triplet Loss for person re-identification in camera sensor networks. Ad Hoc Netw 95
94. Song W, Li S, Chang T, Hao A, Zhao Q, Qin H (2020) Context-Interactive CNN for Person Re-Identification. IEEE Trans Image Process 29:2860–2874
95. Song X, Liu J, Jin Z (2022) Dual Prototype Contrastive learning with Fourier Generalization for Domain Adaptive Person Re-identification. Knowl-Based Syst 256
96. Su C, Zhang S, Xing J, Gao W, Tian Q (2018) Multi-type attributes driven multi-camera person re-identification. Pattern Recogn 75:77–89
97. Subramaniam A, Nambiar A, Mittal A (2019) Co-segmentation inspired attention networks for video-based person re-identification. In: International Conference on Computer Vision, Seoul
98. Sun R, Huang Q, Xia M, Zhang J (2018) Video-Based Person Re-Identification by an End-To-End Learning Architecture with Hybrid Deep Appearance-Temporal Feature. Sens 18(11)
99. Sun R, Lu W, Zhao Y, Zhang J, Kai C (2020) A Novel Method for Person Re-Identification: Conditional Translated Network Based on GANs. IEEE Access 8:3677–3686
100. Tang Y, Yang X, Wang N, Song B, Gao X (2020) Person re-identification with feature pyramid optimization and gradual background suppression. Neural Netw 124:223–232
101. Tao D, Guo Y, Yu B, Pang J, Yu Z (2018) Deep multi-view feature learning for person re-identification. IEEE Trans Circ Syst Video Technol 28(10):2657–2666
102. Tay C-P, Roy S, Yap K-H (2019) AANet: attribute attention network for person re-identifications. In: Conference on Computer Vision and Pattern Recognition, California
103. Tian Y, Li Q, Wang D, Wan B (2019) Robust joint learning network: improved deep representation learning for person re-identification. Multim Tools Applic 78:24187–24203
104. Tian H, Zhang X, Lan L, Luo Z (2019) Person re-identification via adaptive verification loss. Neurocomputing 359:93–101
105. Vishwakarma DK, Upadhyay S (2018) A deep structure of person re-identification using multi-level gaussian models. IEEE Trans Multi-Scale Comput Syst 4(4):513–521
106. Wan C, Wu Y, Tian X, Huang J, Hua X-S (2019) Concentrated local part discovery with fine-grained part representation for person re-identification. IEEE Trans Multim 22(6):1605–1618
107. Wang D, Chen Y, Wang W, Tie Z, Fang X, Ke W (2023) Uncertainty-guided joint attention and contextual relation network for person re-identification. J Vis Commun Image Representation 93
108. Wang G, Lai J, Xie X (2018) P2SNet: Can an Image Match a Video for Person Re-Identification in an End-to-End Way? IEEE Trans Circ Syst Video Technol 28(10):2777–2787
109. Wang H, Hu J (2020) Deep Multi-Task Transfer Network for Cross Domain Person Re-Identification. IEEE Access 8:5339–5348
110. Wang WH, Sun Y, Bi X (2023) Structural redundancy reduction based efficient training for lightweight person re-identification. Information Sciences 637
111. Wang K, Wang H, Liu M, Xing X, Han T (2018) Survey on person re-identification based on deep learning. CAAI Trans Intell Technol 3(4):219–227
112. Wang Z, Jiang J, Wu Y, Ye M, Bai X, Satoh S (2020) Learning Sparse and Identity-Preserved Hidden Attributes for Person Re-Identification. IEEE Trans Image Process 29:2013–2025
113. Wang C, Song L, Wang G, Zhang Q, Wang X (2020) Multi-scale multi-patch person re-identification with exclusivity regularized softmax. Neurocomputing 382:64–70
114. Wang F, Zhang C, Chen S, Ying G, Lv J (2020) Engineering Hand-designed and Deeply-learned features for person Re-identification. Pattern Recogn Lett 130:293–298
115. Wang J, Yuan L, Xu H, Xie G, Wen X (2021) Channel-exchanged feature representations for person re-identification. Inf Sci 562:370–384
116. Wang M, Ma H, Huang Y (2023) Information complementary attention-based multidimension feature learning for person re-identification. Eng Applic Artif Intell 123B
117. Wang P, Zhao Z, Su F, Zhao Y, Wang H, Yang L, Li Y (2020) Deep Multi-Patch Matching Network for Visible Thermal Person Re-Identification. IEEE Trans Multim
118. Wang T, Gong S, Zhu X, Wang S (2014) Person Re-identification by Video Ranking. In: European Conference on Computer Vision, Zurich, Switzerland
119. Wang Y, Wang Z, Jiang M (2019) Part-based pyramid loss for person re-identification. Int J Inf Commun Technol 15(4)
120. Wang Z, Xue J, Wan X, Zhu A, Li Y, Zhu X, Hu F (2022) ASPD-Net: Self-aligned part mask for improving text-based person re-identification with adversarial representation learning. Eng Appli Artif Intell 116
121. Wen Y, Zhang K, Li Z, Qiao Y (2016) A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision. The Netherlands, Amsterdam
122. Wei L, Zhang S, Yao H, Gao W, Tian Q (2019) GLAD: Global–Local-Alignment Descriptor for Scalable Person Re-Identification. IEEE Trans Multim 21(4):986–999

123. Wu A, Zheng W-S, Yu H-X, Gong S, Lai J (2017) RGB-Infrared Cross-Modality Person Re-identification. In: IEEE International Conference on Computer Vision. Venice, Italy
124. Wu L, Wang Y, Gao J, Li X (2018) Deep adaptive feature embedding with local sample distributions for person re-identification. Pattern Recogn 73:275–288
125. Wu L, Wang Y, Li X, Gao J (2018) What-and-where to match: Deep spatially multiplicative integration networks for person re-identification. Pattern Recogn 76:727–738
126. Wu D, Zheng S-J, Bao W-Z, Zhang X-P, Yuan C-A, Huang D-S (2019) A novel deep model with multi-loss and efficient training for person re-identification. Neurocomputing 324:69–75
127. Wu D, Zheng S-J, Zhang X-P, Yuan C-A, Cheng F, Zhao Y, Lin Y-J, Zhao Z-Q, Jiang Y-L, Huang D-S (2019) Deep learning-based methods for person re-identification: A comprehensive review. Neurocomputing 337:354–371
128. Wu D, Zheng S-J, Yuan C-A, Huang D-S (2019) A deep model with combined losses for person re-identification. Cogn Syst Res 54:74–82
129. Wu L, Wang Y, Gao J, Li X (2019) Where-and-When to Look: Deep Siamese Attention Networks for Video-Based Person Re-Identification. IEEE Trans Multim 21(6):1412–1424
130. Wu D, Yang H-W, Huang D-S, Yuan C-A, Qin X, Zhao Y, Zhao X-Y, Sun J-H (2019) Omnidirectional Feature Learning for Person Re-Identification. IEEE Access 7:28402–28411
131. Wu H, Xin M, Fang W, Hu H-M, Hu Z (2019) Multi-Level Feature Network With Multi-Loss for Person Re-Identification. IEEE Access 7:91052–91062
132. Wu L, Wang Y, Shao L, Wang M (2019) 3-D personVLAD: learning deep global representations for video-based person reidentification. IEEE Trans Neural Netw Learn Syst 30(11):3347–3359
133. Wu L, Hong R, Wang Y, Wang M (2020a) Cross-entropy adversarial view adaptation for person re-identification. IEEE Trans Circ Syst Video Technol 30(7):2081–2092
134. Wu L, Wang Y, Yin H, Wang M, Shao L (2020b) Few-shot deep adversarial learning for video-based person re-identification. IEEE Trans Image Process 29:1233–1245
135. Wu Y, Bourahla OEF, Li X, Wu F, Tian Q, Zhou X (2020c) Adaptive graph representation learning for video person re-identification. IEEE Trans Image Process 29:8821–8830
136. Wu Y, Lin Y, Dong X, Yan Y, Ouyang W, Yang Y (2018) Exploit the Unknown Gradually: One-Shot Video-Based Person Re-identification by Stepwise Learning. In: IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA
137. Xiang X, Lv N, Yu Z, Zhai M, Saddik AE (2019) Cross-Modality Person Re-Identification Based on Dual-Path Multi-Branch Network. IEEE Sens J 19(23):11706–11713
138. Xiao T, Li H, Ouyang W, Wang X (2016) Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA
139. Xiong M, Chen D, Chen J, Chen J, Shi B, Liang C, Hu R (2019) Person re-identification with multiple similarity probabilities using deep metric learning for efficient smart security applications. J Parallel Distributed Comput 132:230–241
140. Xu B (2023) Region selection for occluded person re-identification via policy gradient. Image Vis Comput 132
141. Xu B, Liu J, Hou X, Sun K, Qiu G (2019) Cross domain person re-identification with large scale attribute annotated datasets. IEEE Access 7
142. Yan Y, Ni B, Liu J, Yang X (2019) Multi-level attention model for person re-identification. Pattern Recogn Lett 127:156–164
143. Yang F, Yan K, Lu S, Jia H, Xie X, Gao W (2019) Attention driven person re-identification. Pattern Recogn 86:143–155
144. Yang W, Yan Y, Chen S (2019) Adaptive deep metric embeddings for person re-identification under occlusions. Neurocomputing 340:125–132
145. Yao H, Zhang S, Hong R, Zhang Y, Xu C, Tian Q (2019) Deep representation learning with part loss for person re-identification. IEEE Trans Image Process 28(6):2860–2871
146. Yu T, Jin H, Tan W-T, Nahrstedt K (2018) SKEPRID: Pose and Illumination Change-Resistant Skeleton-Based Person Re-Identification. ACM Trans Multim Comput Commun Applic 14(4)
147. Yuan C, Guo J, Feng P, Zhao Z, Luo Y, Xu C, Wang T, Duan K (2019) Learning deep embedding with mini-cluster loss for person re-identification. Multim Tools Applic 78:21145–21166
148. Yuan C, Guo J, Feng P, Zhao Z, Xu C, Wang T, Choe G, Duan K (2019) A jointly learned deep embedding for person re-identification. Neurocomputing 330:127–137
149. Zhang G, Lin W, Chandran AK, Jing X (2023) Complementary networks for person re-identification. Inf Sci 633:70–84
150. Zhang Z, Huang M (2020) Person re-identification based on heterogeneous part-based deep network in camera networks. IEEE Trans Emerg Top Comput Intell 4(1):51–60

151. Zhang J, Hu X, Wang M, Qiao H, Li X, Sun T (2019) Person Re-Identification via Group Symmetry Theory. IEEE Access 7:133686–133693
152. Zhang R, Li J, Sun H, Ge Y, Luo P, Wang X, Lin L (2019) SCAN: Self-and-Collaborative Attention Network for Video Person Re-Identification. IEEE Trans Image Process 28(10):4870–4882
153. Zhang Y, Gu X, Tang J, Cheng K, Tan S (2019) Part-Based Attribute-Aware Network for Person Re-Identification. IEEE Access 7:53585–53595
154. Zhang W, He X, Yu X, Lu W, Zha Z, Tian Q (2020) A Multi-Scale Spatial-Temporal Attention Model for Person Re-Identification in Videos. IEEE Trans Image Process 29:3365–3373
155. Zhang Z, Xie Y, Li D, Zhang W, Tian Q (2020) Learning to Align via Wasserstein for Person Re-Identification. IEEE Trans Image Process 29:7104–7116
156. Zhang L, Li K, Qi Y, Wang F (2021) Local feature extracted by the improved bag of features method for person re-identification. Neurocomputing 458:690–700
157. Zhang R, Fang Y, Song H, Wan F, Fu Y, Kato H, Wu Y (2023) Specialized re-ranking: A novel retrieval-verification framework for cloth changing person re-identification. Pattern Recogn 134
158. Zhang Y, Zhou Z, Li B, Huang Y, Huang J, Chen Z (2019) Improving Slice-Based Model for Person Re-ID with Multi-Level Representation and Triplet-Center Loss. IEICE Trans Inf Syst 102-D(11):2230–2237
159. Zhang Z, Si T (2018) Learning deep features from body and parts for person re-identification in camera networks. EURASIP Journal on Wireless Communications and Networking
160. Zhang Z, Wang Y, Liu S (2020) Cross-domain person re-identification using Dual Generation Learning in camera sensor networks. Ad Hoc Netw 97(102019)
161. Zhao C, Chen K, Wei Z, Chen Y, Miao D, Wang W (2019) Multilevel triplet deep learning model for person re-identification. Pattern Recogn Lett 117:161–168
162. Zheng L, Bie Z, Sun Y, Wang J, Su C, Wang S, Tian Q (2016) MARS: A Video Benchmark for Large-Scale Person Re-Identification. In: European Conference on Computer Vision. Amsterdam, The Netherlands
163. Zheng L, Huang Y, Lu H, Yang Y (2019) Pose-Invariant Embedding for Deep Person Re-Identification. IEEE Trans Image Process 28(9):4500–4509
164. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable Person Re-identification: A Benchmark. In: IEEE International Conference on Computer Vision. Santiago, Chile
165. Zheng W-S, Gong S, Xiang T (2009) Associating groups of people. In: British Machine Vision Conference, London
166. Zheng Z, Zheng L, Yang Y (2019) Pedestrian Alignment Network for Large-scale Person Re-Identification. IEEE Trans Circ Syst Video Technol 29(10):3037–3045
167. Zhong W, Jiang L, Zhang T, Ji J, Xiong H (2019) Combining multilevel feature extraction and multi-loss learning for person re-identification. Neurocomputing 334:68–78
168. Zhong W, Zhang T, Jiang L, Ji J, Zhang Z, Xiong H (2019) Discriminative representation learning for person re-identification via multi-loss training. J Vis Commun Image Represent 62:267–278
169. Zhou S, Wang J, Meng D, Xin X, Li Y, Gong Y, Zheng N (2018) Deep self-paced learning for person re-identification. Pattern Recogn 76:739–751
170. Zhou S, Wang J, Meng D, Liang Y, Gong Y, Zheng N (2019) Discriminative feature learning with foreground attention for person re-identification. IEEE Trans Image Process 28(9):4671–4684
171. Zhou S, Ke M, Luo P (2019) Multi-camera transfer GAN for person re-identification. J Vis Commun Image Represent 59:393–400
172. Zhu F, Kong X, Wu Q, Fu H, Li M (2018) A loss combination based deep model for person re-identification. Multimed Tools and Appl 77
173. Zhu F, Kong X, Zheng L, Fu H, Tian Q (2017) Part-based deep hashing for large-scale person re-identification. IEEE Trans Image Process 26(10):4806–4817
174. Zhu J, Zeng H, Liao S, Lei Z, Cai C, Zheng L (2018) Deep Hybrid Similarity Learning for Person Re-Identification. IEEE Trans Circ Syst Video Technol 28(11):3183–3193
175. Zou G, Fu G, Peng X, Yue L, Gao M, Zheng L (2021) Person re-identification based on metric learning: a survey. Multim Tools Applic 80:26855–26888

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) | 979-8-3503-1590-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACCAI58221.2023.10199860

Deep Learning-based Binary Classification for Spam Detection in SMS Data: Addressing Imbalanced Data with Sampling Techniques

Dr. Manoj Sethi, Naman Tyagi, Parmeet Singh Kalsi and Parupalli Atchuta Rao

Computer Science Department, Delhi Technological University

E-mail : manojsethi@dce.ac.in namantyagi_2k19co241@dtu.ac.in parmeetsinghkalsi_2k19co269@dtu.ac.in parupalliatchutarao_2k19co271@dtu.ac.in

Abstract- This research paper presents a deep learning-based approach for detecting spam in SMS (text) data. The study uses various models namely Dense, LSTM, Bi-LSTM, and GRU to conduct binary classification and predict spam text messages. To address the imbalanced data problem, the study employs undersampling, downsampling, and SMOTE sampling techniques on a public dataset of SMS messages from UCL datasets. The paper presents a study on detecting spam messages in SMS using a dense model. The researchers visualize the commonly used words in spam and non-spam messages and analyze their impact on the model's performance. The findings from this study demonstrate that the proposed dense model exhibits high accuracy in detecting spam messages on the test dataset. This suggests that the model can be useful in identifying spam messages in SMS.

Keywords— Neural Networks; LSTM; Bi-LSTM; GRU; Text Classification; Spam; Ham; Deep Learning.

I. INTRODUCTION

The widespread use of mobile devices and messaging applications has led to an increase in the frequency of unsolicited messages or spam. These spam messages not only cause inconvenience to users but can also pose security threats. Therefore, the development of effective spam detection techniques for SMS text data is crucial.

Spam detection refers to the process of identifying unwanted messages, typically through electronic communication, that are not relevant to the recipient and can potentially cause harm [1], [2]. Different approaches have been proposed for detecting spam, like machine learning-based, statistical and rule-based methods. Machine learning-based approaches, particularly those utilizing deep learning algorithms, have demonstrated significant success in detecting spam messages [3].

Deep learning algorithms are a type of machine learning that takes inspiration from the structure and function of the human brain [4]. They have the ability to learn features from raw data without requiring explicit feature engineering, which makes them more versatile and powerful compared to traditional machine learning algorithms. In the context of natural language processing, some common deep learning algorithms include dense networks, LSTM, Bi-LSTM, and GRU.

The effectiveness of deep learning methods, particularly dense networks, Bi-LSTM, GRU, and LSTM, in natural language processing (NLP) tasks, such as text classification, has been demonstrated in recent years [5], [6]. This research focuses on using these techniques to detect spam in SMS text data. By comparing the performance of these models, the study aims to determine the most effective approach for spam detection.

The proposed approach will involve preprocessing and feature extraction techniques, followed by building and training the deep learning models on the preprocessed data. The performance of the models will be evaluated using quality metrics.

The paper is structured into different sections, and Section 2 specifically provides a review of existing literature on spam detection in SMS text data. Section 3 outlines the Materials and methods used, including the dataset, preprocessing and feature extraction techniques, and deep learning models utilized. Section 4 shows the experimental outcomes and their analysis. Finally, Section 5 winds up the paper and summarizes the research contributions.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

II. Related Work

A multitude of techniques have been suggested for detecting spam in SMS text data, spanning from traditional rule-based methods to machine learning-based approaches [7]. In recent times, deep learning has become increasingly favored in this field due to its capacity to automatically learn intricate features from the input data.

Alvaro Reyes and Juan Garcia proposed an LSTM-based model for SMS spam detection in their work [8]. They preprocessed the data by cleaning and tokenizing the messages, and then used an LSTM network with multiple layers to categorize the SMS as spam or not spam. Rui Ma et al. proposed a hybrid LSTM and CNN model for spam detection in SMS in their work [9]. They used a combination of LSTM and CNN layers to classify the messages after preprocessing the data using tokenization and word embedding techniques.

Suman Saha and Syed S. Islam proposed a Bi-LSTM model for SMS spam detection in their work [10]. They used word embedding approaches to preprocess the data and then used a Bi-LSTM network to categorize the messages as spam or not spam.

Trung and Trang proposed a spam detection system for SMS messages based on a hybrid approach that combines decision trees and SVM [11]. Similarly, Samant and Thakare used random forests to classify SMS messages as spam or non-spam [12].

For instance, Shaikh and Rathod proposed a spam SMS detection model based on convolutional neural networks (CNNs) [13]. Vyas used deep learning techniques for spam SMS detection [14].

Other researchers have also explored more complex deep learning models for this problem. For example, Rahman, Islam, and Akhand proposed a spam SMS message detection system based on a bidirectional LSTM (Bi-LSTM) [15]. Park and Park used gated recurrent units (GRUs) to develop a spam SMS message detection model [16].

Taken together, the proposed techniques display encouraging outcomes in identifying SMS spam, and deep learning-based methods have illustrated their capability to achieve remarkable accuracy in this field. Nonetheless, there is scope for refinement, and more investigation is necessary to examine the effectiveness of alternative deep learning architectures in the detection of SMS spam.

III. Materials and Methods

A. Research Framework

This study utilizes Keras, a Python library for deep learning model development with the TensorFlow backend. Keras provides a range of tools for data preparation, including word tokenization, padding and truncating data, and word embedding [17]. Word tokenization transforms text inputs into sequential data represented as index values of the words. Padding and truncating data ensure that all sequences have the same length, while word embedding converts sequences with multiple dimensions into vectors.

The dense networks, LSTM, Bi-LSTM and GRU algorithms are utilized in order to train the model after the process of data preparation. The models' performances are then evaluated. The focus of this study is to detect spam messages in SMS text data, and the use of models related to deep learning is expected to enhance the precision of detection.

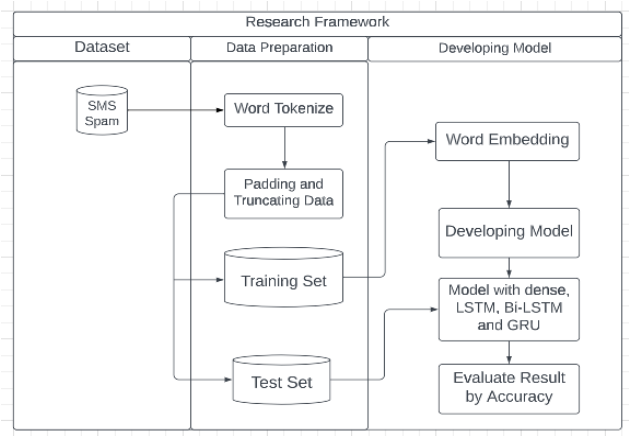


Fig. 1. Research framework.

B. Datasets

The SMS Spam Collection Dataset was selected for this study. The dataset includes 5,574 phone messages that were gathered as part of cellular device spam research and have been categorized as either spam or not spam. This dataset has been extensively used in previous investigations on SMS spam detection and is considered a standard benchmark for this type of research. By utilizing this

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

dataset, the study can ensure that the results obtained are comparable with other studies in the field.

	label	message
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

Fig. 2. Records of dataset for detecting spam.

Dataset for detecting Spam is an imbalanced dataset, with approximately 86% of the messages labeled as ham and 14% labeled as spam. The dataset includes messages in English and contains a variety of spam messages, such as phishing scams, fraudulent offers, and unwanted advertisements.

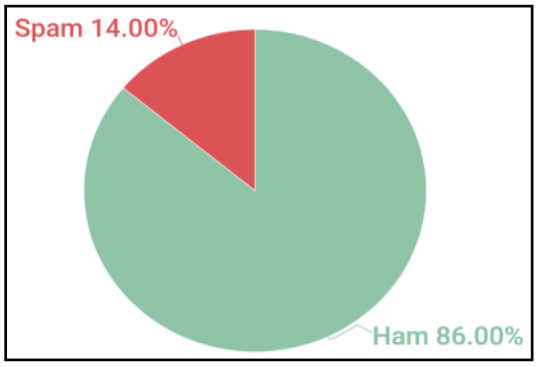


Fig. 3. Pie chart of spam Vs ham.

C. Data Sampling

The dataset consists of 4,825 non-spam (ham) messages and 747 spam messages, indicating an imbalance in the dataset. The most commonly occurring ham message is "Sorry, I'll call later," while the most frequently occurring spam message is "Please call our customer service..." with 30 instances. These results underscore the importance of balancing the dataset prior to training the spam detection model using Dense, LSTM, Bi-LSTM, and GRU architectures.

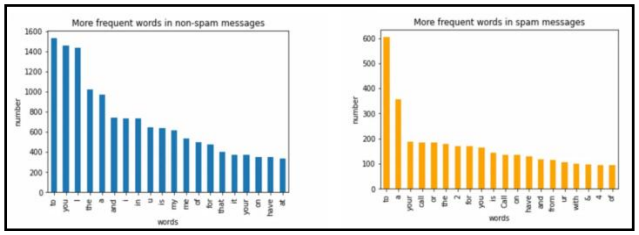


Fig. 4. Most frequent words in spam and ham messages.

The WordCloud for ham messages indicates that the words "now", "work", "How", "Ok", and "Sorry" are frequently used. On the other hand, the WordCloud for spam messages shows that the words "Free", "call", "text", "claim", and "reply" are the most common. The WordClouds provide insights into the vocabulary used in each message type, which can be helpful in developing an effective spam detection model. The bar chart illustrates the class imbalance in the dataset, with ham messages comprising the majority (86%) and spam messages only accounting for a small proportion (14%).

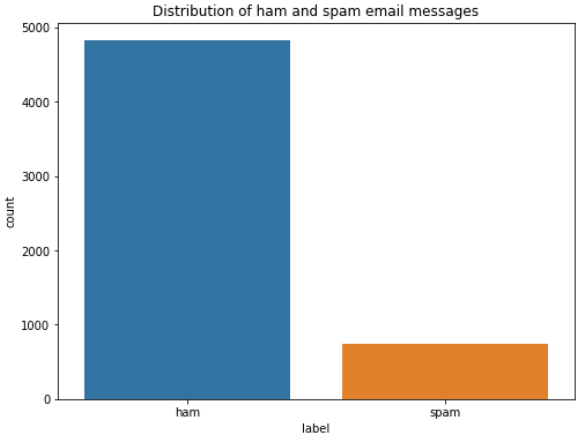


Fig. 5. Distribution of spam and ham messages.

In order to address the imbalanced class distribution in the dataset, various sampling techniques has been employed such as

1. Downsampling : It involves randomly removing instances from the majority class to reduce its size and balance the class distribution with the minority class. This can help prevent the model from becoming biased towards the majority class.
2. *Upsampling :* It involves randomly duplicating instances from the minority class to increase its size and balance the class distribution with the majority class. This can help ensure that the model has enough examples to learn from the minority class.
3. *SMOTE :* A data augmentation strategy called as SMOTE (Synthetic Minority Over-sampling Technique) produces new manufactured instances for the class in deficit by interpolating among existing instances. This technique can assist in increasing the size of the minority class without causing overfitting.

D. Training/Tesing the Data and Text Processing

To prepare the dataset for training deep learning

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

models, the text labels were first converted into numerical numbers. Subsequently, this research divides the data into two subsets, namely, a training and a testing set, in which 80% of the data is allocated for training and the remainder 20% for testing purposes. Additionally, the labels were transformed into numpy arrays to enhance their application in deep learning models.

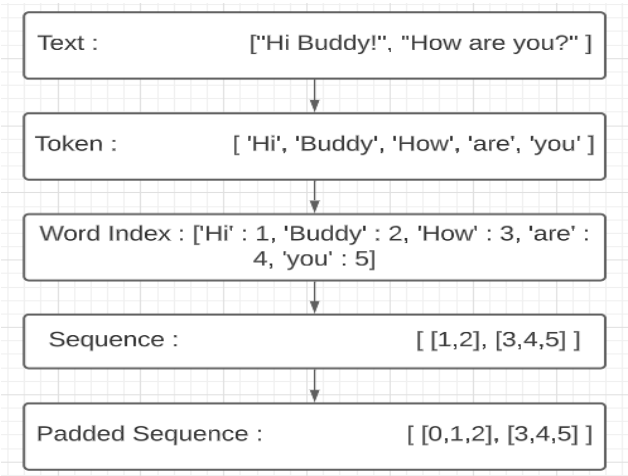


Fig. 6. Process flow in text preprocessing.

Tokenization : Numerical representation of text is essential in deep learning. Tokenization is the initial step in this process. Preprocessing hyperparameters are defined, which includes tokenizing text into words, setting a maximum number of unique tokens to discard rare words, removing punctuation marks, converting all words to lowercase, and converting all words to integer indexes. These hyperparameters will be further explained in the subsequent sections of the paper.

The hyper-parameters used in Tokenizing includes the number of distinctive words to fill in training and testing data, oov_token is utilized in replacing out-of-vocabulary words, and char_level specifies whether to treat every character as a token.

Sequences and Padding : After tokenization, sequences of numbers are obtained to represent each sentence. Then, padding is applied to make all sequences have the same length. Both the types of data : training and testing undergoes through same process.

The padding hyperparameters are used to control the padding process in text sequences. The padding can be added either before or after the sequence using padding = 'pre' or padding = 'post'. The maximum length of sequences can be set using maxlen, where longer sequences will be truncated. The truncating can be done at the beginning or end of the sequences using truncating='pre' or truncating='post'. These hyperparameters play a crucial role in ensuring that all sequences have the same length, which is a requirement for building a neural network model.

E. Modelling

In this study, LSTM, Bi-LSTM, GRU, and dense neural network models were trained and developed for text classification. Each model was evaluated based on its ability to accurately classify spam and non-spam messages.

Dense Model : The dense model is a kind of feedforward neural network. The input layer takes preprocessed text data as input, and the output layer produces a binary classification indicating whether the text is spam or not.

In the training phase, the deep learning model determines the best weights and biases to reduce a loss function that calculates the deviation between the predicted output and the actual output [18], [19]. Binary cross-entropy is the frequently utilized loss function for binary classification. Gradient descent is employed to modify the model's weights and biases. This involves computing the loss function's gradient with respect to the parameters and modifying them in the direction of steepest descent [20].

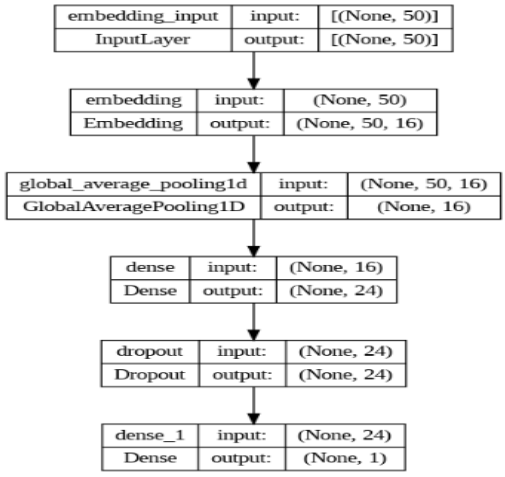


Fig. 7. Model summary of dense model architecture.

The successive model from Keras is utilized to add layers sequentially. The primary layer is the embedding layer, which plots every word to a 16-dimensional real number's vector

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

(embedding_dim). The input layer is defined by input_length = max_len. To minimize the count of parameters and stay away from overfitting, the GlobalAveragePooling1D layer is utilized as the pooling layer. The sigmoid function is appropriate for binary classification tasks. The dense layer and dropout layer are included to prevent overfitting.

The ReLU (Rectified Linear Unit) activation function takes an input value 'x' and returns the maximum value between 0 and 'x'. If 'x' is negative, it returns 0. The sigmoid activation function takes an input value 'x' and gives back a positive less than 1.

To configure and train the model, the 'binary_crossentropy' loss function, 'adam' optimizer, and 'accuracy' metric were utilized in current study. During model fitting process, the training data and labels are passed to the model along with the validation data. The research train the model for 30 epochs and add early stopping functionality via the EarlyStopping() callback function. This stops the training process if the validation loss does not improve following two consecutive epochs, preventing overfitting. Additionally, setting of verbose=2 to display the accuracy and loss on each epoch.

LSTM Model : Natural language processing (NLP) and other sequential data-related apps frequently use recurrent neural networks (RNNs), specifically Long Short-Term Memory (LSTM). LSTM networks have a more intricate architecture than conventional RNNs, which allows them to capture long-term dependencies and prevent the vanishing gradient issue [21].

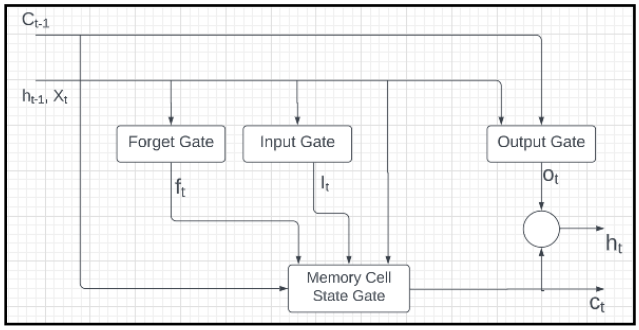


Fig. 8. Structure of LSTM cell.

These networks are built around a memory cell that comprises a hidden state and a cell state. The cell state functions as a conduit of information, carrying it from one point in time to the next, while the hidden state serves as the LSTM's output at each point in time. Three gates—the input gate, the forget gate, and the output gate—are used by LSTMs to alter the status of the cell. The input gate controls how much new information is introduced to the cell state at each time step, and the forget gate controls how much information is subtracted [22]. Finally, the output gate controls the portion of the cell state that is utilized to compute the output. LSTMs are highly desirable for their ability to prevent the problem of vanishing gradient, which commonly affects conventional RNNs.

embedding_1_input	input:	[(None, 50)]
InputLayer	output:	[(None, 50)]
embedding_1	input:	(None, 50)
Embedding	output:	(None, 50, 16)
spatial_dropout1d	input:	(None, 50, 16)
SpatialDropout1D	output:	(None, 50, 16)
lstm	input:	(None, 50, 16)
LSTM	output:	(None, 20)
dropout_1	input:	(None, 20)
Dropout	output:	(None, 20)
dense_2	input:	(None, 20)
Dense	output:	(None, 1)

Fig. 9. Model summary of LSTM model architecture.

Bi – LSTM : BiLSTM builds on the LSTM architecture by processing the input sequence in both forward and backward directions using two separate LSTM layers. This makes it possible for the model to incorporate data from the input sequence's past and future. In a BiLSTM architecture, each input element is processed by both the forward and backward LSTMs, and the outputs from these two LSTMs are concatenated. This concatenated output is then fed into the next layer of the network, which can be another BiLSTM layer or a fully connected layer for classification or regression tasks.

Gated Recurrent Unit (GRU) : Another variety of recurrent neural network includes GRU model, that is similar to LSTM but with less parameters. It uses two gates, namely the reset gate and the update gate, to manage the stream of information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

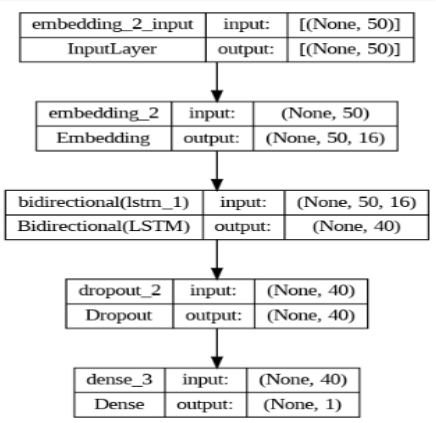


Fig. 10. Model summary of Bi-LSTM model architecture.

The reset gate ensures how much of the previous hidden state is retained for the current time step. It is a vector that is multiplied element-wise with the previous hidden state to determine which parts of it to forget and which parts to keep [23]. This mechanism allows the model to selectively remember or forget past information as needed.

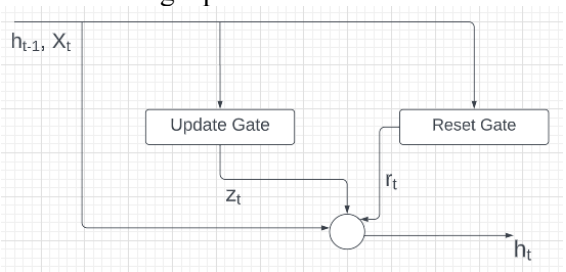


Fig. 11. Structure of GRU cell.

The update gate also ensures how much of the existing input and last hidden state are used to modify the memory cell. It is a vector that is computed based on the existing input and last hidden state, and helps in deciding the amount of information that should be added to the cell.

The hidden gate is the output of the GRU at each time step, and it is computed based on the existing input and last hidden state [24]. It is also used as the input to the next time step, allowing the model to capture information from both the past and the present of the input sequence.

GRUs have the benefit of handling long-term dependencies in sequential data without experiencing the vanishing gradient issue.

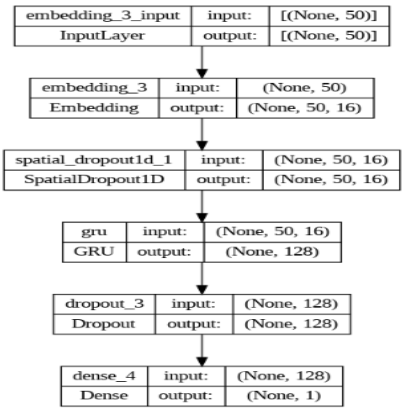
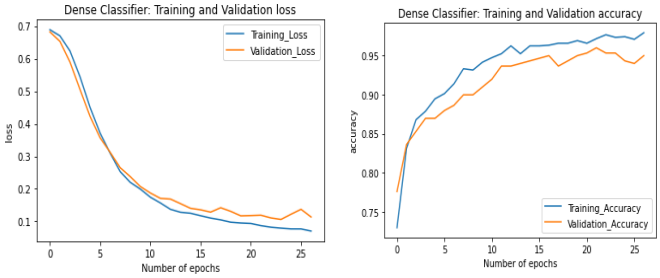


Fig. 12. Model summary of GRU model architecture.

V. Results and Discussions

After training the model, the study evaluated the performance of dense network, LSTM, Bi-LSTM and GRU models and plotted the graph of loss and accuracy by number of epochs. The research also compares the results of models with different sampling techniques such as downsampling, upsampling and SMOTE sampling. However, the accuracy of model based Dense and Bi-LSTM is similar but dense model outperformed other models in different sampling techniques.

Dense Model

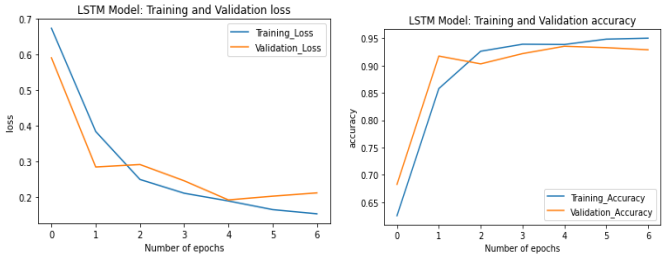


(i). (ii).

Fig. 13. Training and Validation plots in dense model (i) loss vs epochs (ii) accuracy vs epochs.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

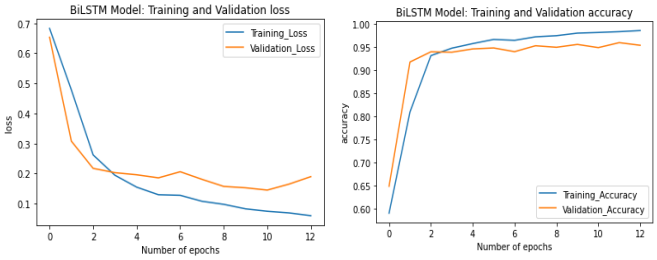
LSTM Model



(i). (ii).

Fig. 14. Training and Validation plots in LSTM model (i) loss vs epochs (ii) accuracy vs epochs.

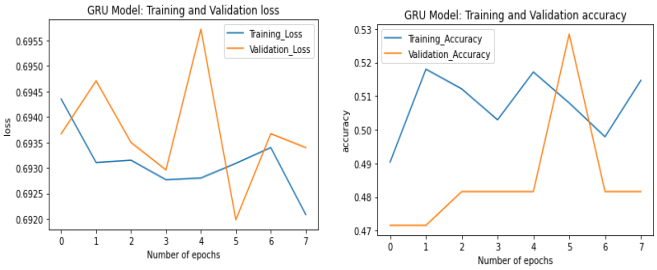
Bi-LSTM Model



(i). (ii).

Fig. 15. Training and Validation plots in Bi-LSTM model (i) loss vs epochs (ii) accuracy vs epochs.

GRU Model



(i). (ii).

Fig. 16. Training and Validation plots in GRU model (i) loss vs epochs (ii) accuracy vs epochs.

For the Dense model, the best results are obtained using upsampling, with the lowest loss and highest accuracy achieved. This indicates that upsampling was able to effectively balance the dataset and improve the performance of the model.

TABLE I. Comparison of Deep Learning Models with Different Sampling Techniques.

Models	Resampling Techniques					
	DownSampling		UpSampling		SMOTE	
	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
Dense	**0.082**	**0.9665**	**0.036**	**0.9891**	**0.11**	**0.9715**
LSTM	0.173	0.9531	0.092	0.9782	0.11	0.9689
BiLSTM	0.142	0.9565	0.062	0.9870	0.05	0.9829
GRU	0.693	0.4816	0.060	0.9777	0.06	0.9849

For the LSTM and BiLSTM models, the best results are obtained using upsampling and SMOTE techniques, with similar performance between the two techniques. This suggests that both techniques were able to address the class imbalance issue and improve the execution of the models.

In contrast, for the GRU model, the best results are

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

obtained using upsampling and SMOTE techniques, with the highest accuracy achieved using SMOTE. This indicates that both techniques were effective in improving the performance of the model, which was otherwise performing poorly with down sampling.

Overall, the results suggest that different resampling techniques can have varying effects on the performance of deep learning models, and the best technique will depend on the specific dataset and problem being addressed.

VI. Conclusion

In this research, the selection of the resampling technique for spam classification tasks using deep learning models can have a considerable influence on their performance. The performance of the models varied depending on the sampling technique used, with upsampling, downsampling, and SMOTE techniques each showing different levels of effectiveness depending on the specific model and dataset.

Overall, upsampling appeared to be the most effective technique, consistently improving the performance of the models across all architectures, while downsampling showed poor performance in all cases. SMOTE was also shown to be effective, particularly for the LSTM and BiLSTM models, but its effectiveness varied depending on the specific model being used.

These findings suggest that researchers and practitioners working on spam classification tasks should carefully evaluate the performance of different resampling techniques to identify the best method for their specific dataset and model.

Overall, the use of resampling techniques can be an effective strategy for addressing class imbalance and improving the performance of deep learning models for spam classification tasks.

References

[1] K. V. Amulya, K. R. Mohan and M. S. Jaiswal, "SMS Spam Detection Using Deep Learning Techniques: A Comparative Study," in International Journal of Advanced Science and Technology, vol. 30, no. 6, pp.167-178, 2021.

[2] S. Singh and S. Agrawal, "SMS Spam Detection using Deep Learning Techniques: A Comparative Study," International Journal of Intelli- gent Systems and Applications, vol. 12, no. 3, pp. 1-9, 2020, doi: 10.5815/ijisa.2020.03.01.

[3] A. Hassanpour and M. R. Khayyambashi, "A Survey of SMS Spam Detection Techniques using Machine Learning and Deep Learning Approaches," Journal of Network and Computer Applications, vol. 168, pp. 102764, 2020, doi: 10.1016/j.jnca.2020.102764.

[4] X. Li, Y. Wang, and S. Lai, "An Improved SMS Spam Detection Algorithm Based on Deep Learning," Journal of Intelligent & Fuzzy Systems, vol. 37, no. 6, pp. 7727-7737, 2019, doi: 10.3233/JIFS-179507.

[5] S. R. Goudar, M. H. M. Khan and S. S. Suri, "SMS Spam Detection Using Deep Learning and Multi- Objective Optimization," in Interna- tional Journal of Machine Learning and Cybernetics, vol. 12, no. 6, pp. 1511- 1522, 2021. doi: 10.1007/s13042-020-01208-3.

[6] O. A. Omotosho, J. Akinyelu and B. O. Fatile, "SMS Spam Detection using a Hybrid Deep Learning Architecture," 2022 IEEE 16th Interna- tional Conference on Intelligent Systems and Control (ISCO), pp. 82-87, Jan. 2022, doi: 10.1109/ISCO53562.2022.9620396.

[7] H. Hassan, F. Akram, and S. M. Mir, "A Hybrid Feature Engineering and Deep Learning Approach for SMS Spam Detection," Computers & Electrical Engineering, vol. 99, pp. 106993, 2022, doi: 10.1016/j.compeleceng.2022.106993.

[8] A. Reyes and J. Garcia, "SMS Spam Detection using LSTM," IEEE Access, vol. 6, pp. 23350-23359, 2018.

[9] R. Ma, P. Zhou, and J. Peng, "A Hybrid LSTM and CNN Model for Spam Detection in SMS," IEEE Access, vol. 7, pp. 30939-30947, 2019.

[10] S. Saha and S. Islam, "SMS Spam Detection Using BiLSTM and Word Embedding," International Journal of Computer Science and Network Security, vol. 19, no. 1, pp. 169-177, 2019.

[11] T. N. Trung and N. N. T. Trang, "A hybrid approach for spam detection in SMS messages," in Proc. 2019 International Conference on Advanced Computing and Applications, pp. 220-225, 2019.

[12] V. S. Samant and V. M. Thakare, "SMS spam filtering using random forests," in Proc. 2019 International Conference on Computing and Communication Technologies for Smart Nation, pp. 1-6, 2019.

[13] C. M. Shaikh and N. M. Rathod, "Spam SMS detection using convo- lutional neural networks," in Proc. 2020 International Conference on Innovative Computing and Communication, pp. 180-184, 2020.

[14] S. S. Vyas, "Spam SMS detection using deep learning," in Proc. 2020 5th International Conference on Computing, Communication and Security (ICCCS), pp. 1-5, 2020.

[15] M. A. Rahman, M. R. Islam, and M. A. H. Akhand, "Spam SMS Message detection using bidirectional LSTM," in Proc. 2021 IEEE Region 10 Symposium (TENSYMP), pp. 163-166, 2021.

[16] S. H. Park and J. H. Park, "Spam SMS message detection using gated recurrent unit networks," in Proc. 2022 14th International Conference on Advanced Communication Technology (ICACT), pp. 577-581, 2022.

[17] O. A. Omotosho, J. Akinyelu, and B. O. Fatile, "Effective SMS Spam Detection using Attention-based Hybrid Deep Learning Approach," Expert Systems with Applications, vol. 184, pp. 115511, 2022, doi: 10.1016/j.eswa.2021.115511.

[18] J. Akinyelu, B. O. Fatile and O. A. Omotosho, "Effective SMS Spam Detection using Deep Learning Techniques and Feature Selection," 2022 IEEE 16th International Conference on Intelligent Systems and Control (ISCO), pp. 75-81, Jan. 2022, doi: 10.1109/ISCO53562.2022.9620386.

[19] J. Akinyelu, B. O. Fatile, and O. A. Omotosho, "SMS Spam Detection using Deep Learning with Feature Selection based on Mutual Informa- tion," IEEE Access, vol. 9, pp. 26766-26777, 2021, doi: 10.1109/AC-CESS.2021.3069585.

[20] S. Biswas and M. S. Hossain, "SMS Spam Detection Using Deep Learning and Ensemble Techniques," in IEEE Access, vol. 9, pp. 16246- 16257, 2021. doi: 10.1109/ACCESS.2021.3051591.

[21] N. R. Swamy and A. C. Reddy, "An Efficient SMS Spam

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

Detection Model Using Deep Learning," in IEEE Access, vol. 9, pp. 112844- 112856, 2021. doi: 10.1109/ACCESS.2021.3100877.

[22] S. Singh and S. Agrawal, "SMS Spam Detection using Recurrent Neural Network and Attention Mechanism," 2021 International Conference on Computer Communication and Informatics (ICCCI), pp. 1-4, Jan. 2021, doi: 10.1109/ICCCI50389.2021.9371469.

[23] M. Sharma and A. Kumar, "SMS Spam Detection using Convolu- tional Neural Network with Attention Mechanism," 2021 11th Inter- national Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 244-249, Apr. 2021, doi: 10.1109/CONFLU-ENCE52351.2021.9454796.

[24] A. M. Alghamdi, M. Al-Mulhem and M. Al-Rakhami, "SMS Spam Detection Using Deep Learning," in IEEE Access, vol. 8, pp. 136251-136262, 2020. doi: 10.1109/ACCESS.2020.3002663

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:26:30 UTC from IEEE Xplore. Restrictions apply.

Design and Control of a 2kW Electric Vehicle Charger

Sushant Chaturvedi
Electrical Engineering
Delhi Technological University
Delhi, India
sushantchaturvedi017@gmail.com

Alka Singh
Electrical Engineering
Delhi Technological University
Delhi, India
alkasingh.eed@gmail.com

2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES) | 979-8-3503-4729-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSSES58299.2023.10201070

***Abstract*—An interleaved boost converter-based full bridge phase shift converter (PSFB) for charging an electric vehicle is illustrated in this paper. To attain a PF around unity at the AC input of supply, the suggested charger uses an interleaved boost converter as a front-ended device that runs in continuous conduction mode (CCM), and PSFB to process the battery charging of the EV. Almost all boost converters run at a fixed frequency with pulse width modulation (PWM), and the PSFB converter tries to maintain the power required by the battery. This paper gives a systematic and detailed design for a 2.0 kW prototype along with a controller design for the said converters, which is designed and validated with simulation using MATLAB/Simulink whose results show that at 100 kHz switching frequency, peak efficiency is 95%, THD is maintained to be less than 5% in variation from half load to full load, and PF is constant at 0.99 from 1950 W to 2.2 kW.**

Keywords— EV charger, phase-shift full-bridge converter (PSFB), DC-DC converter, power factor correction (PFC), total harmonic distortion (THD), interleaved boost converter (ILBC).

I. INTRODUCTION

Non-renewable energy sources and their supply are rapidly running out.; Thus, research has focused more and more on renewable or "sustainable" energy sources. Electric vehicles (EVs) are gaining popularity in the automotive sector. EVs are mainly powered by motors powered by battery, and hence best possible battery technology is a requirement for now, to make E-Vehicles a proper replacement to in use conventional IC engine vehicles, it is essential to develop a battery charging device working on high power with good efficiency, excellent power factor.

As the market in Electric Vehicle (EV) sector expands, so does the demand for quick and efficient battery charging services. To reach the demand, research in the field of battery technology attempts to push the boundaries of quick charging on several fronts, such as the type of materials used, incorporated topologies and design of cell package. Although from the perspective of energy transfer, the vehicle battery and its charger are the most pivotal elements of the charging process, Nevertheless, the adopted charging technique also has a big impact on the entire procedure, specifying how this circuital energy gets transmitted from the charger to the vehicle battery, to make EVs a viable choice. [1]

In today's light electric vehicles (LEVs) sector, the power capacity of the main battery is kept at around 10kW. At this rating, a 2 kW battery charger is decent to use. With the ability to be plugged directly into an AC mains outlet, on□board chargers are handy, but regulating the supply of current power from sources like domestic, commercial, and charging stations is important.[2]

Typically, a diode bridge rectifier (DBR) followed by a DC link capacitor is initially constituted in battery chargers, which is followed by an isolated DC-DC converter working on a high frequency. This arrangement leads to poorly formed and out of phase current waveform to a voltage waveform, producing a low input power factor which adversely affects the supply, and to match the required quality standards, a power factor correction (PFC) device is also used in the battery charger. [3]

The objective here is to develop an efficient, safe, economical better power quality EV Charger with low charging time. Out of various acceptable line current conditioning converters and PFC topologies are explained in [4-6], the ILBC has shown to be one of the most dependable and efficient solutions for line regulation as well for supply current shaping. And similarly among all available different topologies of DC-DC converter [7-9], The PSFB converter as it has a high efficiency among all, becomes a recommended DC-DC converter for car batteries. This is made possible by the PSFB converter devices' ability to switch at zero voltage, which allows for a greater operating frequency and reduced magnetic size for the converter..

II. TOPOLOGY SELECTION

In Fig. 1 the employed EV Charger topology diagram. On the primary side of the system, we have an interleaved boost converter. The interleaved boost converter has the advantage of paralleling the switching devices. Cancelling the ripple (likely to be reduced to half at least) at the output while effectively increasing the switching frequency, it also reduces stress on the output capacitor, hence reducing the filter size. As the wattage rating rises, it is frequently necessary to connect converters in a cascaded manner. Interleaving two boost converters is frequently used in high-power applications to enhance performance and decrease the size of the PFC device. Furthermore, for higher current applications and voltage step-ups, the currents flowing through the switches become mere fractions of the supply current. [4]

For the n-channel interleaved PFC circuit, the PFC channels are operated (360/n) phase shifted from each other. The study in [6] shows that to achieve high efficiency and PF the interleaved configuration is preferred. The results of different simulations in Table 1 show the efficiency for interleaved topology is over 96% and the PF can be up to 0.99 for the entire universal line range at full load.

The second stage of the charger is a DC-to-DC converter. It regulates the charging and discharging of the vehicle, here on the secondary side, regulation is achieved with the phase shift full bridge rectifier (PSFB) maintaining the power flow control to the battery.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:12 UTC from IEEE Xplore. Restrictions apply.

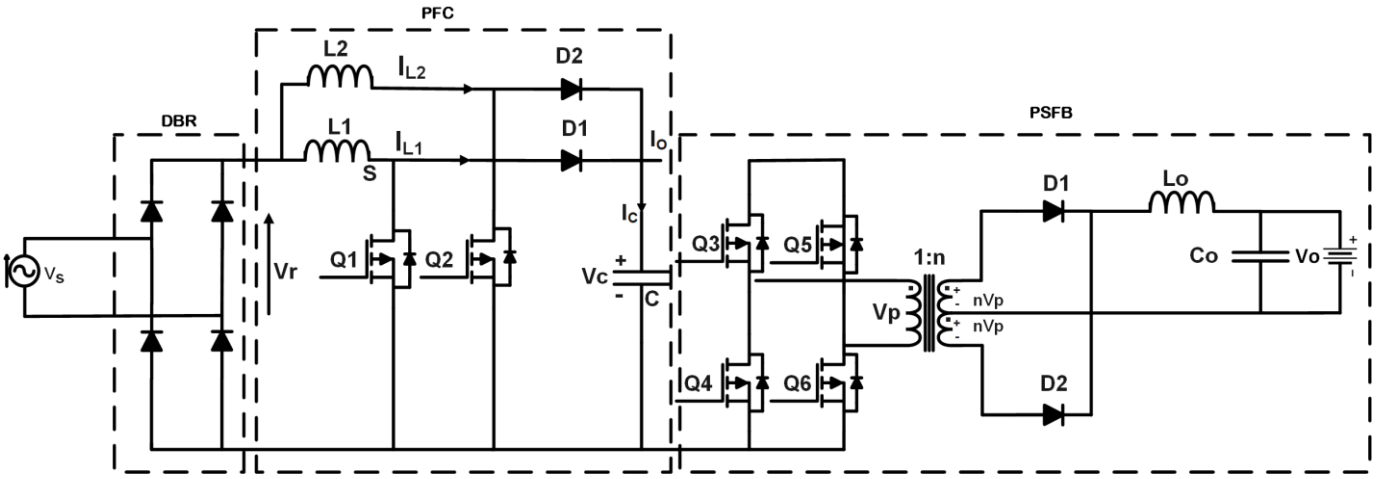


Figure 1 : Circuit Diagram of a proposed PFC Interleaved boost converter fed PSFB converter.

Table 1: COMPARISON OF AC–DC PFC TOPOLOGIES FOR BATTERY CHARGING

Topology	Efficiency	THD	Power Rating
Boost	93	5.64	<1000W
Interleaved	97	3.4	<3000W
Bridgeless	96	4.83	>3000W
Semi-Bridgeless	98	5.0	>3000W

III. WORKING PRINCIPLE OF PROPOSED CHARGER

The said EV charger device as shown in Fig. 1 employs two separate converters, each with a specific set of functions as explained in the prior segment. Though the foremost aim of the PFC interleaved converter is to keep supply current distortion and harmonics free with a well-regulated DC link voltage, The PFC output voltage, V_c, acts as an input of the PSFB converter, which charges the EV battery. As the regulation in the line voltage and so in DC link voltage is less likely.

An interleaved dual inductor boost PFC converter is proposed for the first-stage setup. So it regulates the output voltage to the desired value while improving the input power factor.The Average current controller is used to achieve PFC. Interleaved PFC topology corporate power switches and their body diodes for the rectification process. This topology operates with either duty ratio D higher than 0.5 or lower than 0.5 depending on whether the input voltage is less than or greater than half of the output voltage. As mentioned in [10], Interleaving is achieved by a phase shift of 360/n° between the PWM signals of the switch controllers. Two current loops are shown in Fig. 2(a) is provided for two inductors L_1 and L_2. The input current is the sum of the two inductor currents in L_1 and L_2. The ripple currents among these interleaved inductors are out of phase, which makes them neutralize the effect, therefore bringing down the high-frequency ripple current of supply caused by the converters' switching action, and so the EMI filter at the input can be made smaller. For equal sharing of load between converter interleaved channels, the net reference value of voltage loop controller is chopped down to half, which is given as a reference to every current controller.

Working of a PSFB DC-DC converter as explained in detail in [11,12]. Switching legs operate within fewer than 50% of duty, but they're given a shift in phase between both the switching legs of the converter. This phase shift determines enhancement in voltage, The switching pulses are separated by dead time, given to the switching leg in order to give parasitic capacitance adequate time to release or store the charge. The output variation of this converter is influenced by the duty cycle and its loss. There can be multiple sources for duty cycle loss in this type of converter. One among them can be because to the dead time given between the switching legs, and the other can be due to ΔD which occurs with the reversal of the primary current. Marking secondary voltage to zero, Effective duty ratio (D_e) is given by

$$D_e = D - \mu - \Delta D \tag{1}$$

Where μ is the value of loss in the duty cycle due to dead time. Indecent value of dead time can result incorporate in the loss of ZVS, reduced efficiency, or even up to the device failure due to shoot-through.

IV. DESIGN OF PROPOSED CHARGER MODEL

The design of the converter is carried out for a 2 kW operation rating, working in CCM mode at the switching frequency in kHz's with assumptions such as that the switching devices are ideal is considered for ease of understanding.

A. The PFC InterleavedBoost Converter

A 230V AC (RMS) supply with a supply frequency of 50Hz is given to the filter circuit which uses a capacitor as a filter. Once after rectification, capacitors remove the ripple and smoothen the pulsating DC output so a near to constant DC voltage is supplied. For high power application, more than one boost cell can be operated in parallel to increase the power capability of the PFC circuit and reduce the capacitor current ripple, known as interleaved system.

For our system to work at the rated output power of 2kW, with the regulated DC link voltage (V_c) maintain at 400V under normal conditions. The duty cycle for the same can be expressed as

$$D = \frac{V_C - V_i}{V_C} \tag{2}$$

where V_i is the peak value of the voltage supply and can be represented as $\sqrt{2}$ times of AC line Voltage (V_S). At times of the lowest AC line voltage supply, i.e 200V the maximum current from the input line will pass and so will the highest value of Duty Cycle will be reached which will make the (2) as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:12 UTC from IEEE Xplore. Restrictions apply.

$$D_{\max} = \frac{V_C - \sqrt{2}\,V_{S\min}}{V_C} = \frac{400 - \sqrt{2}(200)}{400} = 0.3 \quad (3)$$

Considering the inductor ripple current to be 20% of the inductor current for both of the inductor values.

$$\begin{aligned}\Delta i_{L1} = \Delta i_{L2} &= \frac{P_o}{V_C}.20\% \\ &= \frac{2000}{400} * 0.20 \\ &= 1A\end{aligned} \quad (4)$$

Major part of design process is the selection of inductor value, as that assures the proper PFC functionality. The selection of this value depends on inductor ripple current.

$$L = \frac{V_i * D}{f_S . \Delta i_L} \quad (5)$$

From (3), the value of inductor L_1 *and* L_2 are calculated as

$$\begin{aligned}L_1 = L_2 &= \frac{400 * 0.3}{50000 * 1} \\ &= 2.4mH\end{aligned} \quad (6)$$

The ripple in DC link voltage (ΔV_C) is taken at 1% of its nominal value.

$$\Delta V_C = \frac{1 * 400}{100} = 4V \quad (7)$$

From (6) the output capacitor can be estimated at the given voltage ripple.

$$\begin{aligned}C &\geq \frac{P_O}{4\pi f_{Supply} * V_C * \Delta V_C} \\ &\geq 1986\mu F\end{aligned} \quad (8)$$

B. The PSFB DC-DC Conveter

In this part of the system, voltage at the DC link capacitor output of the previous stage converter works as input of an isolated full bridge phase shift (PSFB) DC to DC Converter for the vehicle charging. Batteries of these vehicles are generally made of lithium ion phosphate (LFP) cell, charging around at 48V.

The equation for output voltage in PSFB can be expressed as,

$$V_O = 2 * V_C * n * D \quad (9)$$

where, V_O is the demand voltage of battery load from the converter for charging, V_C is the input fed to the DC-DC Converter from the previous state and n is the ratio of secondary to primary side of transformer and D is the duty cycle for the converter.

The Value of D should be less than 50% for avoiding the saturation of the core [11].

$$n = \frac{Ns}{Np} = \frac{50}{400} = 0.125 \quad (10)$$

If the output Voltage of the charger is expected to be at 48V the from (9) & (10) we get D to be 48%,

$$\begin{aligned}48 &= 2 * 400 * 0.125 * D \\ D &= 0.48\end{aligned} \quad (11)$$

The Value of Filter Inductor is calculated for 20% inductor ripple current,

$$\begin{aligned}\Delta i_{Lo} &= \frac{P_O}{V_O}.20\% \\ &= \frac{2000}{48} * 0.2 \\ &= 8.3A\end{aligned} \quad (12)$$

Transformer leakage inductance (L_k) often referred as shim inductance also, calculated with equation (13) as in the same load condition, it stores more energy than parasitic capacitance does, helping in achieving the Zero Voltage Switching (ZVS) condition.

$$L_K > 2C_{OAVG(eff)} X \frac{V^2_{dc}}{I_P} \quad (13)$$

Hence, form (12) we get the value of output inductor

$$L_O = \frac{V_O(1-D)}{2 * \Delta i_{LO} * f_{SW}} \quad (14)$$

The switching frequency (f_{SW}) of the DC-DC Converter is 500 KHz, upon substituting the parameters in (14) we get,

$$\begin{aligned}L_O &= \frac{48(1-0.48)}{2 * 8.3 * 100000} \\ &= 15.03\mu H\end{aligned} \quad (15)$$

Similarly, the output capacitor depends upon the voltage ripple and can be measured by,

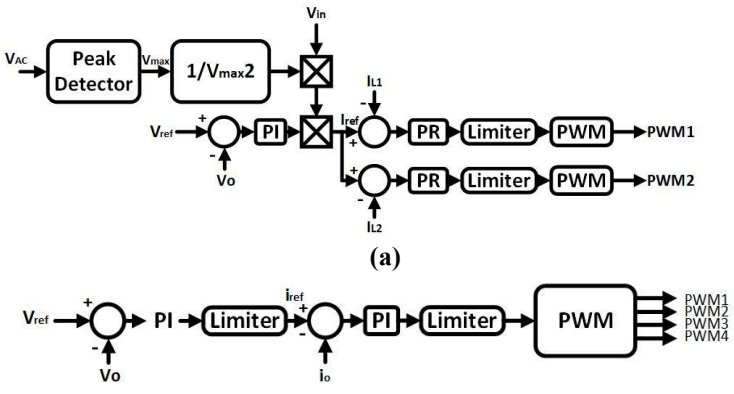


Figure 2: Control Algorithm for (a) PFC Converter (b) PSFB DC-DC Converter

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:12 UTC from IEEE Xplore. Restrictions apply.

$$C_O = \Delta_{iL} * \frac{D}{f_{SW} * \Delta V_O} \quad (16)$$

Where the ΔV_O is the output ripple voltage, kept at 1%, giving the output capacitance to be around 100uF.

V. CONTROL SCHEME OF PROPOSED SYSTEM

Controller of both the PFC part & PSFB Converter part of the charger, shown below in fig. 2(a) and (b) respectively are made of two controller loops, one for a Voltage and other for Current Control. Small signal modeling is crucial to the right controller design for any converter controller. The parameters for the controllers' are derived from the required dynamic behavior, requirements, & aims [5], [9]. The interleaved bridgeless power stage is almost identical to the traditional boost converter. As a result, the converter's single channel is modeled.

A. *Design of Voltage Control Loop*

A voltage control loop is a pre-regulator with a sluggish response time, here we measure the output voltage and compare it with a reference voltage, the difference in both the voltages is what we consider as an error and that is fed to an PI controller, the output of which is given to a PWM generator which generates the Phase shift PWM signal pairs, which are fed to the switching devices.

B. *Design of Current Control Loop*

Current control loop is immediate in nature, and creates inductor currents so that they are sinusoidal and in-phase with the rectified line voltage input. For all current control loop reference current (i_L*) is obtained via outer band of voltage control loop. This Controllers' Output is compared with an saw tooth carrier signal with unity amplitude, which generates the duty cycle. Since the reference current is sinusoidal in shape, PR controller gives better performance and zero steady state error

In PSFB DC-DC Converter a limiter is provided at both the voltage and current control loop of the controller, while the limiter at voltage loop restricts the battery charging current and the same limits the phase shift of PWM to 180°.

VI. SIMULATION RESULT

A simulation prototype of the suggested EV charger is simulated using the MATLAB/SIMULINK software. Table 2 lists all of the parameters that were determined in section IV.

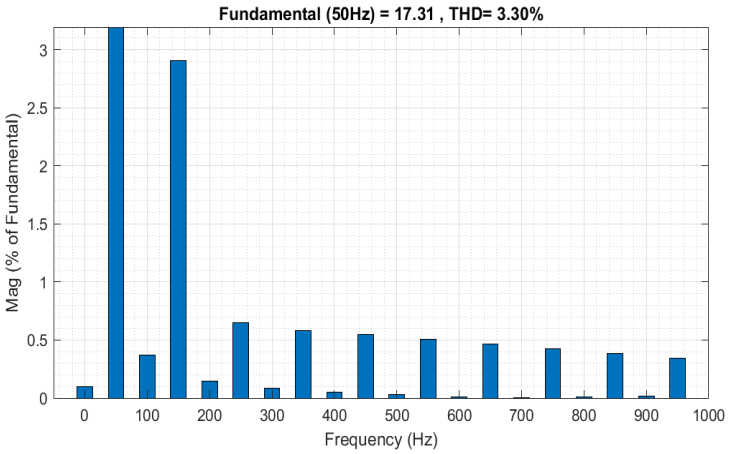


Figure 3: THD (%) of the input current

Table 2: PARAMETERS FOR TAKEN BATTERY CHARGING MODEL

Parameter	Variable	Value
Supply Voltage	V_s	230V
Input inductor	L_1,L_2	3mH
DC-link Voltage	V_C	400V
DC-link Capacitance	C	2000 µ F
Transformer Ratio	n	0.125
Filter Inductance	L_o	15 µ H
Filter Capacitance	C_o	100 µ *F*
Output Power	P_o	2kW
Switching Frequency	F_{sw}	100Khz

The performance results of both the stages of the charger are displayed and discussed under steady and dynamic behavior .

A. Performance at steady-state for the PFC-Interleaved Boost converter

Under steady-state operation, Waveforms of the voltage supplied from grid, supply current and the voltage at the output of the PFC converter are provided in Fig.5 for the said test conditions at switching frequency of 50 kHz the output of the bus voltage is maintained at 400Vdc. The supply current and the supply voltage are well in phase, having a nearly perfect sinusoidal waveform, as expected. Maximum amplitude of the sinusoidal waveform of the AC mains supply current is scaled up 5 times in waveform for sinusoidal behavior study. With a maximum inductor current of 6.5A, I_{L1}, I_{L2} are functioning in CCM which is effectively obtained using the Proportional-Integral Controller.

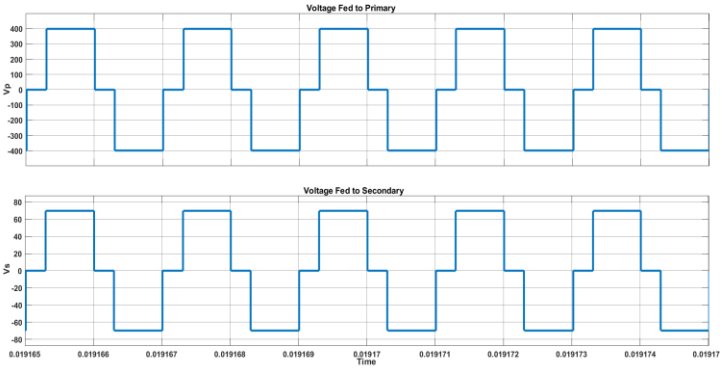


Figure 4: Primary and Secondary Voltages Waveform of high frequency transformer

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:12 UTC from IEEE Xplore. Restrictions apply.

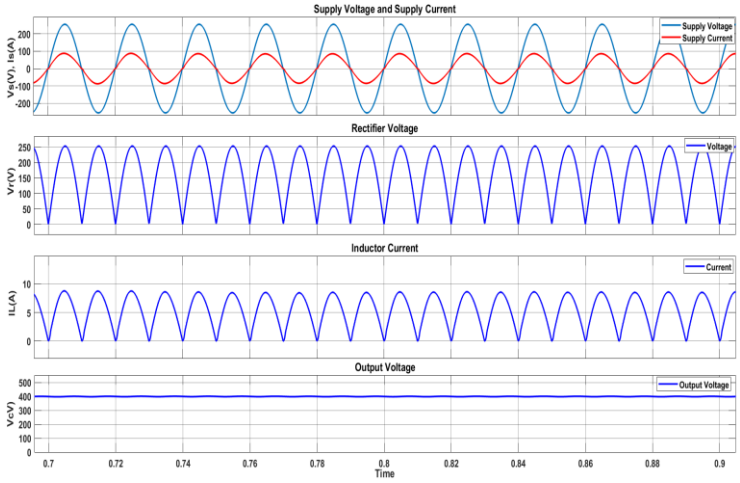


Figure 5: Steady-State performance for supply current (Is) and Voltage (Vs), Rectified Voltage (V_r), Inductor current (I_{L1}, I_{L2}) & DC-link voltage (V_{dc}) at supply voltage (Vs) = 220V

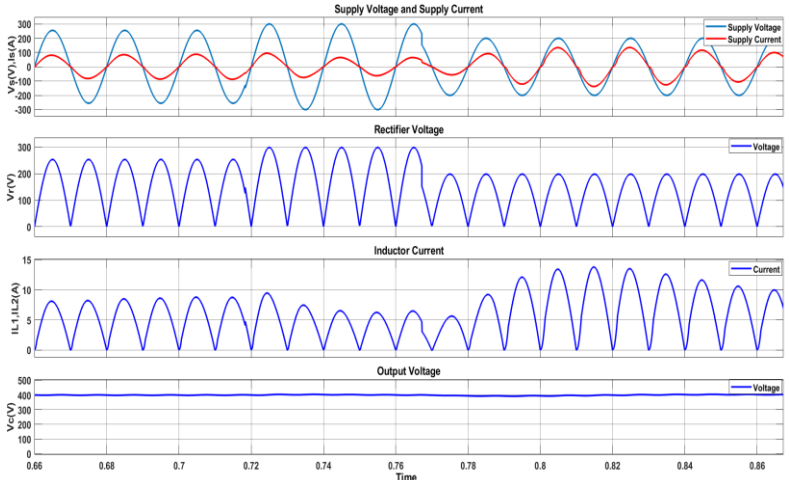


Figure 7: System Dynamic performance of supply current (Is) and Voltage (Vs), Rectified Voltage (V_r), Inductor current (I_{L1}, I_{L2}) & DC-link voltage (V_{dc}) at variable supply voltage.

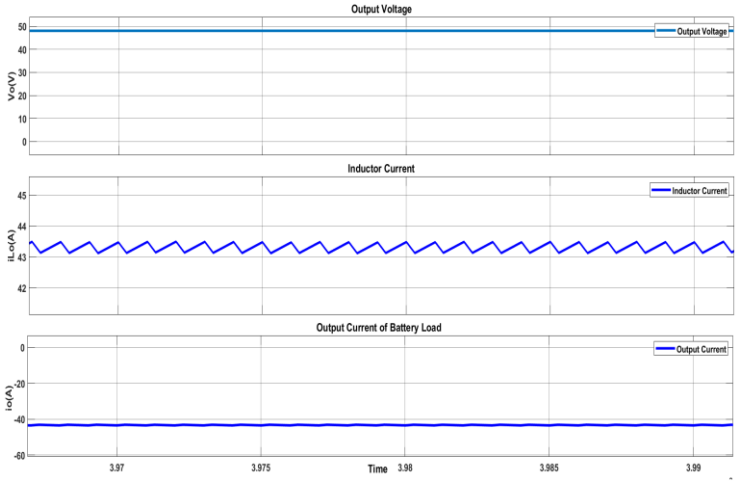


Figure 6: Steady-State performance of PSFB DC-DC for Output Voltage (V_O), Inductor current (I_{LO}) & Battery Load Current (i_O) at full load

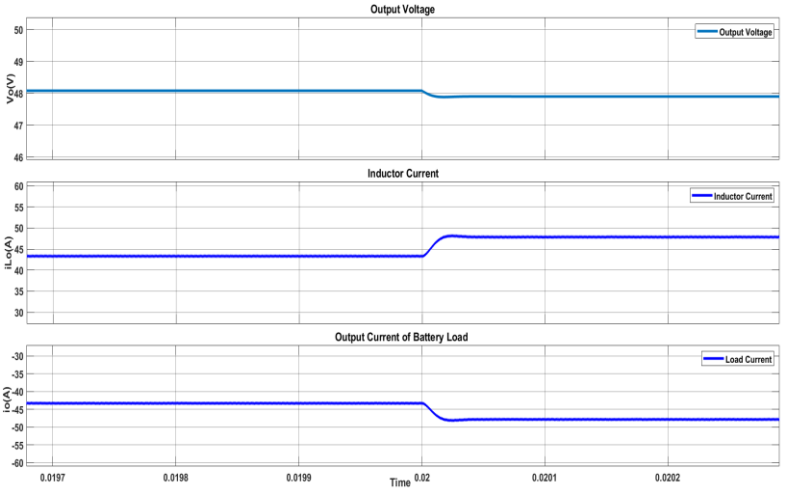


Figure 8: System dynamic performance of PSFB DC-DC for Battery Voltage (V_O), Inductor current (I_{LO}) & Battery Load Current (i_O) at variable load

B. Performance at steady-state for the PSFB converter.

The phase shifted waveforms of primary and secondary voltages of high frequency transformer are shown in Fig. 6. Given a load current of 42A, the output of PSFB converter is regulated at 48V. Inductor current has a 43.5A peak value when the current waveform is triangular, which is the region most suited for the converter's ideal performance, as seen by the waveforms of I_{Lo}.

C. Performance of the dynamic behavior for the PFC-Interleaved Boost converter.

Fig.7 displays dynamic performance of PFC-Boost converter over a variety of input voltages. To examine the dynamic behavior, PFC-Boost is analyzed with varying the supply voltage to 10% up and down.In the beginning, V_s was kept at 230V initially as in ideal condition and after that rises to 250V at 0.7 second which then later dropped to 200V at time 0.77 second. Figure 7 also shows that the source current decreases as the supply voltage rises and rises as the supply voltage falls with a harmonic free waveform. I_{L1} & I_{L2} both are following same patter as shown in fig. 7 and. V_{dc} is maintained to be constant throughout even with varying V_s because of good PI-controller tuning.

D. Performance of the dynamic behavior for the PSFB converter.

The second stage converter's dynamic performance is confirmed by changing load from 2.0KW to 2.20KW. PI-Controller well regulates and keeps the output voltage constant. Io goes from 44A to 48A in terms of load current. Figure 8 illustrates the same behavior of the load variation of the system.

VII. CONCLUSION

Design of an EV battery charger rated at 2KW, which consists of ILBL type PFC as a front end converter followed by a PSFB converter is described in this paper. Power factor of the supply is obtained near to unity. Results of the simulated model are presented with design model and waveforms. This topology has less switching losses, better PF value and can also be implemented for high power ratings, exhibiting excellent transient and steady state behaviors. The system PF range i also provided for the supply voltage variation ranging from 180 V to 250 V input. The power factor is maintained at 0.99 for full load. The designed converter gains its peak efficiency of 95% at 100kHz switching frequency, 220V input and 2 kW output power.

Table 3 summarizes information for various Vs and the resulting PFs, Is, THD and system efficiency at constant 2KW full load.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:12 UTC from IEEE Xplore. Restrictions apply.

Table 3: SYSTEM CURRENT THD AND SYSTEM EFFICIENCY WITH RESPECT TO SUPPLY VOLTAGE

V_S (V)`	I_S(A)	PF	THD (%)	Efficiency (%)
180	11.7	0.995	3.46	95.7
190	11.5	0.996	4.22	92.2
200	10.75	0.992	3.56	93.7
210	10.2	0.9994	4.02	94.2
220	9.6	0.9998	4.12	95.0
230	9.34	0.9998	3.47	94.6
240	9.0	0.9998	3.19	94.4
250	8.5	0.9997	3.51	94.8

VIII. FUTURE SCOPE

The future scope EV chargers largely depends on the development and adoption of electric vehicles (EVs) and the associated infrastructure. While Level 1 chargers are relatively slower compared to higher-level chargers, they still play a significant role in the overall EV charging landscape. Residential level charging, workplace and destination charging, smart and wireless charging are some potential future trends and possibilities for Level 1 EV chargers.

REFERENCES

[1] Utsav Sharma, and Bhim Singh, "An Onboard Charger for Light Electric Vehicles," 2020 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES).

[2] Ionel "Dan" Jitaru, Nicolae Daniel Bolohan, "A High Efficiency 2KW DC-DC Converter for Automotive Application," 2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC).

[3] Rahul Pandey and Bhim Singh, "A Power Factor Corrected Electric Vehicle Battery Charger Using Boost Converter 2018 8th IEEE India International Conference on Power Electronics (IICPE).

[4] João Paulo M. Figueiredo and Fernando L. Tofoli, "A Review of Single-Phase PFC Topologies Based on The Boost Converter," 2010 9th IEEE/IAS International Conference on Industry Applications - INDUSCON 2010.

[5] Murat Yilmaz and Philip T. Krein, "Review of Battery Charger Topologies, Charging Power Levels, and Infrastructure for Plug-In Electric and Hybrid Vehicles," IEEE TRANSACTIONS ON POWER ELECTRONICS, VOL. 28, NO. 5, MAY 2013.

[6] Hyeun-Tae Cho, Jingook Kim, Jeehoon Jung and Katherine A. Kim, "Comparison of Input Power Factor Correction Techniques for Buck Converters in Single-Phase Wireless Power Transfer Systems," 2015 IEEE PELS Workshop on Emerging Technologies: Wireless Power (2015 WoW).

[7] Chan-Song Lee, Jin-Beom Jeong, Baek-Haeng Lee, and Jin Hur, "Study on 1.5 kW Battery Chargers for Neighborhood Electric Vehicles", 2011 IEEE Vehicle Power and Propulsion Conference.

[8] Dingsihao Lyu, Thiago Batista Soeiro, and Pavol Bauer, "Impacts of Different Charging Strategies on the Electric Vehicle Battery Charger Circuit Using Phase-Shift Full-Bridge Converter", 2021 IEEE 19th International Power Electronics and Motion Control Conference (PEMC).

[9] Mohd Shahnawaz Khan, Shelas Sathyan, Harinaik Sugali,v and S S Chandra Bommagani," Design of On-Board Battery Charger using Interleaved Bridgeless Type PFC and Phase Shifted Full Bridge Converter",2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science.

[10] Fariborz Musavi, Wilson Eberle, and William G. Dunford, "A High-Performance Single-Phase Bridgeless Interleaved PFC Converter for Plug-in Hybrid Electric Vehicle Battery Chargers",IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS, VOL. 47, NO. 4, JULY/AUGUST 2011.

[11] Arya P S, and Chithra R., "Phase Shifted Full Bridge DC-DC Converter", International Research Journal of Engineering and Technology (IRJET).

[12] Marian K Kazimierczuk, "Pulse width Modulated DC DC Power Converters, Second Edition".

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:12 UTC from IEEE Xplore. Restrictions apply.

Connection Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ccos20

Design of advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks

Gebrekiros Gebreyesus Gebremariam, J. Panda & S. Indu

To cite this article: Gebrekiros Gebreyesus Gebremariam, J. Panda & S. Indu (2023) Design of advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks, Connection Science, 35:1, 2246703, DOI: 10.1080/09540091.2023.2246703

To link to this article: https://doi.org/10.1080/09540091.2023.2246703

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 22 Aug 2023.

Submit your article to this journal

Article views: 98

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ccos20

CONNECTION SCIENCE
2023, VOL. 35, NO. 1, 2246703
https://doi.org/10.1080/09540091.2023.2246703

OPEN ACCESS Check for updates

Design of advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks

Gebrekiros Gebreyesus Gebremariam[a,b], J. Panda[b] and S. Indu[b]

[a]Department of Electronics and Communication Engineering, Raya University, Tigray, Ethiopia; [b]Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

ABSTRACT

Wireless sensor networks (WSNs) are an emerging military and civilian technology that uses sensors. Sensor networks are hierarchical and chaotic in remote, unmonitored sites. Wireless sensor networks pose unique security threats due to their public location and wireless transmission. WSNs are vulnerable to various routing attacks, including Black holes, Sybil, sinkholes and wormholes. In this paper, we proposed advanced intrusion detection systems based on hybrid machine learning (AIDS-HML) in wireless sensor networks to identify and classify attacks. Hybrid machine learning classifiers identify wireless sensor network dangers. Benchmark datasets are used to compare the proposed model to baseline models in terms of precision, recall, f1-score, and accuracy. The scheme is trained and evaluates prediction models. This confirms that the detection accuracy achieved 99.80% using the NSL-KDD benchmark dataset based on hybrid random forest and extreme gradient boost (RF-XGB). The hybrid cluster labelling K-Means (CLK-M) s achieved better classification accuracy of 100% using UNSW_NB15, and CICIDS2017 benchmark datasets for binary classification of label attacks. Different attack detection metrics were compared against various benchmark datasets to evaluate the quality of this work. The proposed system is efficient in simulations for feature extraction and route discovery and detection attacks achieving an accuracy of 99.46%.

ARTICLE HISTORY
Received 2 May 2023
Accepted 5 August 2023

KEYWORDS
Hybrid security technique; intrusion detection system; wireless sensor networks; malicious nodes; detection and classification; hybrid machine learning models

1. Introduction

Wireless sensor networks (WSNs) are Affordable and low-power sensor nodes (Gao, 2014). Sensor nodes with multi-hop routing and self-organising intelligent sensor networks fall into this category (Gebreyesus, 2021). WSNs are autonomous spatially disseminated devices using sensors for physical and environmental conditions. Cooperatively, nodes in WSNs collect data on some physical or environmental characteristic, including noise, vibration, temperature, pressure, motion, or pollution, and relay that data to a hub node (Alsaedi et al., 2017; Farooqi et al., 2013; Han et al., 2019; Pundir et al., 2020). Raw data sensed and collected

CONTACT Gebrekiros Gebreyesus Gebremariam kiros2004comp@gmail.com Department of Electronics and Communication Engineering, Raya University, Maychew 7020, Tigray, Ethiopia Department of Electronics and Communication Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bwana Road, 110042, Delhi, India

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

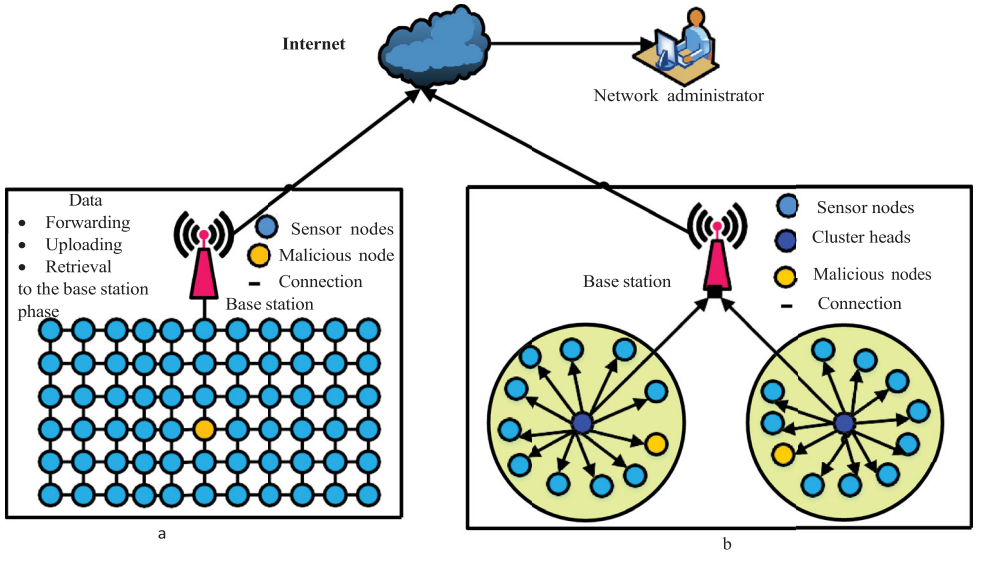


Figure 1. Architecture of distributed (a) and hierarchical (b) WSNs.

sent to the cluster head was stored and analysed at the base station (Blywis, 2009; Choi et al., 2004) as shown in Figure 1. WSNs are composed of self- configured and connected by radio signals having a low operating battery and low cost distributed hierarchically and randomly (Saravana Kumar et al., 2021; Singh et al., 2017). WSNs are recent technology and have gained significant attention for research scenarios (Singh et al., 2017). They comprise low-power and cost sensors randomly distributed over the target localisation. The sensors are distributed to act on specific tasks (Mangrulkar & Negandhi, 2018; Rouissi et al., 2019). The sensors have sensing and signal processing capabilities and activate wireless communication in WSNs (Vinitha et al., 2019). In WSNs, A gat way provides wireless connectivity using wired and distributed nodes. WSNs are the main components of the internet of things (IoT), with small, low-power sensors for data collecting and monitoring from the environment (Saidi et al., 2020; Zhao et al., 2023). Wireless sensor nodes can be configured as distributed or hierarchical fashion in the target area for deployment as shown in Figure 1.

The frequency of cyber-attacks targeting international businesses is increasing, leading to the rapid development of intrusion detection systems (IDS) in both industry and academia (Mahbooba et al., 2021). The availability and confidentiality of the data may have been compromised as a result of attempts by attackers to break the network's security through vulnerabilities in the security measures (Abdulganiyu et al., 2023). IDS is a network security solution that collects and analyses network data to detect abnormal behaviour and protect system resources (Zhang et al., 2020). It is crucial in maintaining network security (Wang et al., 2021). Anomalies and misuse are two types of intrusions that can occur in WSNs. Anomaly detection utilises mathematical models and compares estimated feature values with reference values to identify deviations from normal behaviour (Godala & Vaddella, 2020). Misuse detection, on the other hand, relies on previously observed malicious activities and their specific patterns to identify intrusions.

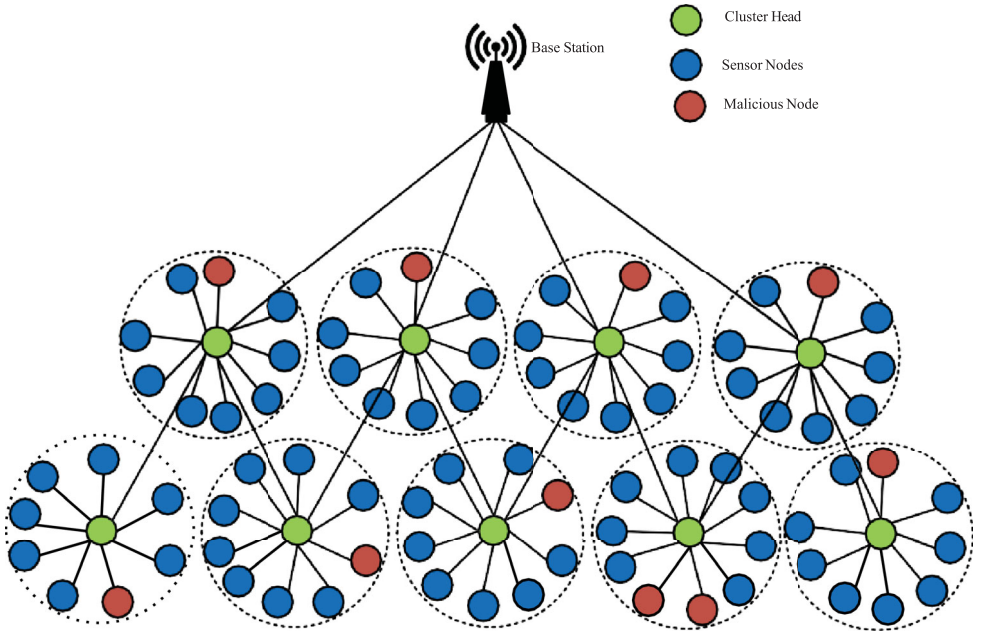


Figure 2. Wireless sensor networks model with clustering Nodes for data aggregation to the sink node.

Common causes of cybercrimes include DoS and web attacks. Attacks in WSNs can be active or passive, involving unauthorised eavesdropping, information gathering, and packet manipulation (Elsaid & Albatati, 2020). To ensure protection, WSNs employ two layers of defence. intrusion detection and prevention technologies. IDS tracks intruders, their activities, time, location, and network layer, providing valuable information to system administrators (Elsaid & Albatati, 2020). Cybercrimes pose significant threats to businesses as they introduce malicious attacks into networks. Intrusion detection is vital to cybersecurity and involves examining and identifying security breaches in an information system.

Machine learning (ML) techniques offer generic solutions and continuously improve their performance (Praveen Kumar et al., 2019). In WSNs, ML finds applications in various fields, enhancing performance and reducing the need for manual maintenance. It facilitates information extraction from large sensor-generated data, enabling machine-to-machine (M2M) communications, cyber-physical systems, and the Internet of Things (IoT). ML proves valuable in WSNs for optimising sensor nodes, energy management, localisation, node classification, routing, and detecting attacks across multiple layers of sensor networks. Machine learning strategies are employed for selecting the cluster head to enhance energy performance in Wireless Sensor Networks (WSNs) as shown in Figure 2.

ML techniques also helpful in detecting and removing malicious nodes at the cluster head, improving network reliability and operational lifetime (Praveen Kumar et al., 2019). ML algorithms efficiently classify DoS attacks and handle data aggregation and forwarding tasks to the sink node. Data aggregation plays a critical role in WSNs, impacting energy consumption, storage requirements, network load, and processing speed as in Figure 2. Various data aggregation methods, such as cluster-based, tree-based, in-network,

or centralised approaches, can improve network reliability and lifespan. ML algorithms offer advantages in collecting and organising data, contributing to the security and efficiency of WSNs.

1.1. Problem statement

The problem addressed in this research is developing an advanced Intrusion Detection System (IDS) based on hybrid machine learning (AIDS-HML) to protect hierarchically structured Wireless Sensor Networks (WSNs). Current IDSs face difficulties achieving high detection accuracy and minimising false alarms. The hierarchical organisation of sensor nodes adds complexity to the design of advanced IDSs (Paliwal & Kumar, 2018).. The objective is to create a robust and efficient IDS that effectively identifies and responds to intrusions while considering the hierarchical structure of the WSN. Specific challenges to be addressed include:

1. Developing a hierarchical IDS architecture that can adapt to the structure of the WSN and effectively handle data collection and aggregation.
2. Extracting relevant features from sensor data in a resource-constrained environment and employing techniques to reduce dimensionality and select the most informative features.
3. Integrating multiple machine learning algorithms in a hybrid approach to leverage their strengths and enhance intrusion detection accuracy.
4. Addressing the challenge of stolen sensitive information in WSNs, where the absence of physical defence lines makes it crucial to monitor network traffic flow.
5. Evaluating the proposed IDS in terms of detection accuracy, recall, precision, false alarm rate, resource utilisation, and scalability.

Addressing these challenges will contribute to developing advanced IDSs tailored for hierarchically structured WSNs, providing enhanced security and maintaining the efficiency of the network. The proposed solution will aim to overcome the limitations of existing IDSs and provide insights into the effective integration of hybrid machine learning techniques in the context of WSNs. The NSL-KDD, UNSW_NB15, and CICIDS2017 datasets are utilised for training and testing, serving as benchmarks for evaluating the proposed system. Specifically, the NSL-KDD dataset, which includes four types of attacks (Root to Local attacks, Denial of Service, Probing, and User to Root attacks), is used to test the effectiveness of the proposed approach.

1.2. Research motivation

Attacks affect Wireless Sensor Networks (WSNs) applied in different applications, including healthcare systems, industrial applications, traffic light management systems, and intelligent power generation and distribution systems for keeping the real-time demand. So it is essential to survey and study hybrid security methods by combining two schemes, such as hybrid IDS, hybrid routing, and hybrid anomaly detection techniques. Hybrid security techniques are famous for the detection of attacks in WSNs. The hybrid security techniques utilise datasets and hybrid models to evaluate the system's effectiveness for attack

classification and detection. The significant contributions of this work are to explore the various hybrid security techniques in wireless sensor networks as follows.

1. Design and simulate a secure attack detection scheme against routing attack scenarios in wireless sensor networks.
2. Explore the various design of hybrid security techniques using a combined attack-defence strategy.
3. Explore data processing transformation techniques for evaluating security performance using benchmark datasets.
4. Exploration of the various hybrid machine learning models for effective DoS attack detection and classification using a public dataset to evaluate the system's effectiveness.
5. Explore wormhole tunnels and routing techniques to provide the optimal path for data transmission nodes and suitable solutions for secure routing and monitoring mechanisms.
6. Analyse the Location and detection of multiple attacks like a wormhole and black hole attacks using hybrid and combined schemes.
7. Explore the hybrid techniques used for WSNs traffic analysis and detecting complicated attacks with the public dataset with normal and malicious behaviours of the network traffic data.

1.3. Paper of contribution

This paper's primary contribution is to set a novel, hybrid machine-learning approach for intrusion detection to bolster the security and performance of WSNs while minimising resource consumption. In the past, many IDS strategies have been implemented with benchmark machine learning approaches to improve the detection accuracy of WSN assaults. However, as the quantity and interconnectivity of wireless sensor nodes grow, the frequency with which routing assaults are launched also rises. Here are some key contributions of this research:

- Design and planning of secure advanced intrusion detection systems for scalable and resources optimised WSNs.
- The research specifically addresses the challenges of hierarchically structured WSNs. By considering the multi-level organisation of sensor nodes, the IDS can adapt its detection mechanisms more effectively, providing improved coverage and response throughout the network.
- Explore how advanced hybrid machine learning models can be used in WSNs intrusion detection systems with metrics to examine the most important aspects of intrusion detection systems.
- The research provides valuable insights into the effective integration of different machine learning approaches in the context of WSNs. This knowledge can be extended to other domains where hybrid ML solutions may be beneficial.
- It helps decrease the number of false alarms and boosts the efficiency of intrusion detection systems for Secure data transmission across the network for data privacy and protection in WSNs.

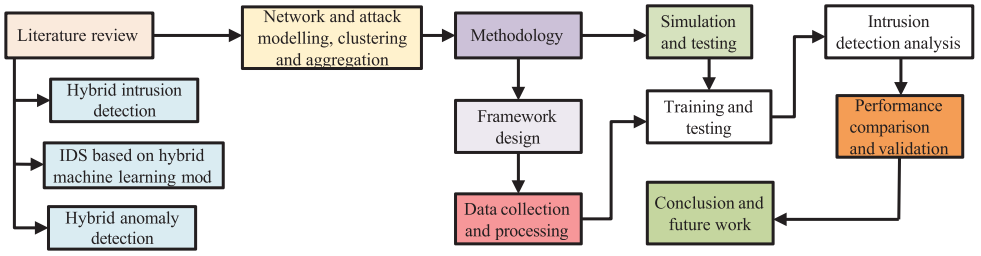


Figure 3. The organisation and framework of the proposed work.

The other sections of this paper are divided into distinct parts for easier readability and comprehension. The first section provides a high-level overview of wireless sensor networks, sensor clustering, our inspiration for this study, and the results. Section 2 summarises prior research on sophisticated machine-learning strategies that utilise IDS data. Section 3 explains the network models underpinning the proposed hybrid security system for intrusion detection. Methodology and dataset processing for hybrid security methods are described in Section 4. In part 5, we will examine the simulation and its outcomes. Section 6 depicts the experimental setup, data collection, and analysis, while the final section discusses the takeaways and suggestions for the future as depicted as shown in Figure 3.

2. Related works

This section's extensive literature review is conducted on various hybrid security techniques, including hybrid intrusion detection systems, machine learning techniques, and anomaly detection using benchmark datasets. The hybrid techniques effectively detect WSN attacks such as wormholes, flooding, sinkhole, and Sybil attacks in WSNs. These include hybrid intrusion detection systems, advanced HIDS, and intelligent and artificial immune hybrid intrusion detection systems (AIHIDS) (Singh et al., 2017). The hybrid routing protocols, hybrid misuse, hybrid optimisation algorithms, anomaly detection techniques, and hybrid clustering techniques are discussed in this section.

2.1. Hybrid intrusion detection systems

(Singh et al., 2017) presented an Advanced Hybrid Intrusion Detection System (AHIDS) for Wireless Sensor Networks (WSNs). AHIDS employs a cluster-based architecture with an improved LEACH protocol to reduce power consumption in sensor nodes. Hybrid Artificial Neural Networks (HANNs) are used to detect and classify potential threats like Sybil, wormhole, and Hello flood attacks. The proposed system achieved high detection rates: 99.40% for Sybil attacks, 98.20% for Hello flood attacks, and 99.20% for wormhole attacks. Similarly, (Cepheli et al., 2016) examined a hybrid intrusion detection technique that combines flexible and tunable parameters using parallel detection methods. Their hybrid system, guided by a central node, improves DDoS attack detection accuracy. Figure 4 shows signature-based and anomaly-detection block diagrams for attack detection as normal traffic and DDoS attacks.

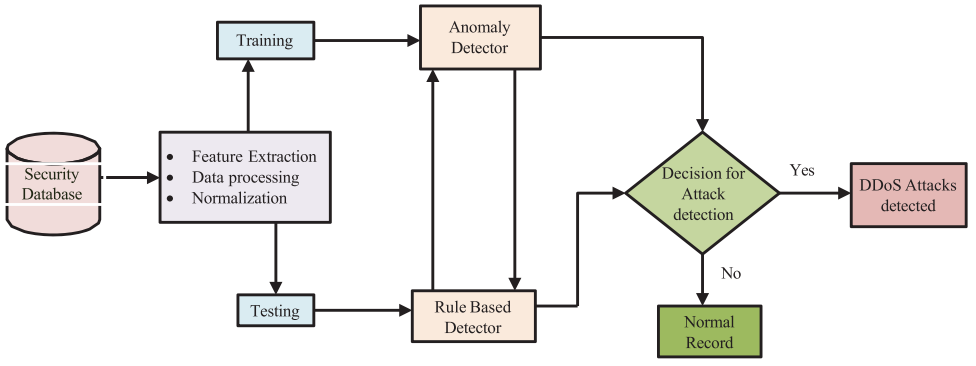


Figure 4. hybrid intrusion detection model with anomaly detector and rule-based detector.

The detection process analyses network traffic and extracts building activity model features. DARPA 2000 dataset is used to evaluate the intrusion detection system, focusing on DDoS attacks and normal network traffic. The hybrid detection employs anomaly and signature detectors. Anomaly detector identifies normal and abnormal traffic data through feature extraction, while the signature detector uses predefined sets for traffic data features. AHIDS uses anomaly detection blocks to identify normal and abnormal data packets and misuse detection blocks to recognise various attacks as shown in Figure 4. The detection of malicious nodes in AHIDS, based on fuzzy rules, involves three steps as follow:

- It measures the transmission of data packet history in WSNs through base nodes.
- It selects the feature set and looking the key elements for packet classification.
- Anomaly intrusion detection techniques are established based on data packet resolution.

The fuzzy-based intrusion detection technique uses MPNN, which consists of BPNN and FFNN for anomaly and misuse detection, as shown in Figure 5. It is applied for the highest detection rate using supervising learning technique. The fuzzy base AHIDS with FFNN and BPNN achieves greater attack detection accuracy using massive clustered training. The multilayer perception is utilised for estimating the error rate, e_i, using the formula as in Equation (1).

$$e_i = d_i - a_i \tag{1}$$

Where d_i represents the preferred output and a_i is the true output obtained from MPNN. The MPNN model consists of BPNN and FFNN and is applied to evaluate the detection accuracy of the various class attacks in WSNs.

The MPNN utilises BPNN and FFNN techniques for IHIDS to manage huge datasets and the system's stability. FFNN detects the new type of attacks, and BPNN clusters the mysterious attacks for MPNN supervised learning. The membership vector applied for the fuzziness F(V) is given by Equation (2).

$$F(V) = -\frac{1}{n}\sum_{k=1}^{n}(\mu_i \log \mu_i) + (1-\mu_i)\log(1-\mu_i) \tag{2}$$

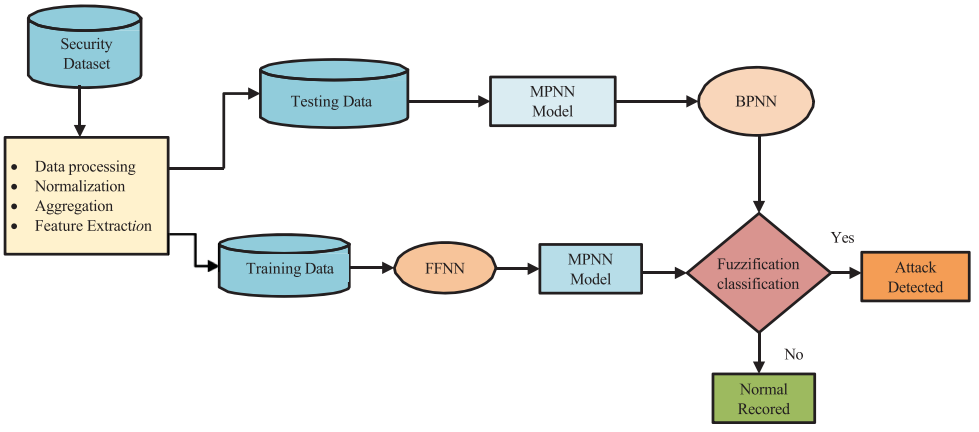


Figure 5. Block diagram for advanced and hybrid intrusion detection system.

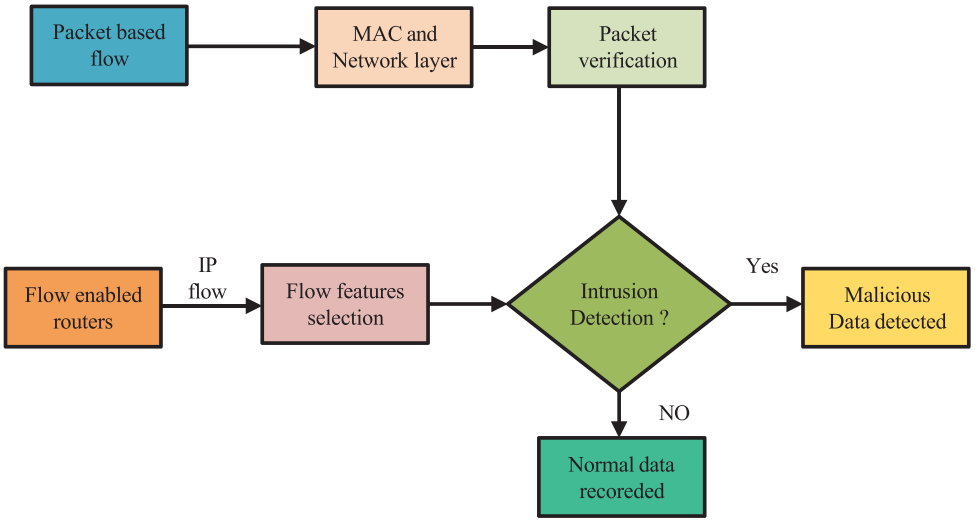


Figure 6. Block diagram of an intrusion detection system using the flow-based technique.

Where $V = \{\mu_1, \mu_2, \ldots, \mu_n\}$ is a set of fuzzy, the fuzziness values are categorised into high, low, and mid fuzziness groups with training and testing samples.

(Gandhimathi & Murugaboopathi, 2020) conducted research on flow-based and cross-layer hybrid intrusion detection. Their approach aims to detect anomaly traffic and identify potential attacks based on narrow features. The process involves two phases: In the first phase, flow-based IDS is used to classify malicious nodes, and in the second phase, packets are analysed using cross-layer features to verify and detect potential threats. The flow-based IDS monitors network traffic, utilising network information to classify data as either normal or malicious, as shown below in Figure 6. Flow-based anomaly detection can exhibit a regular profile of behaviour by keeping tabs on network activity and keeping track of many parameters.

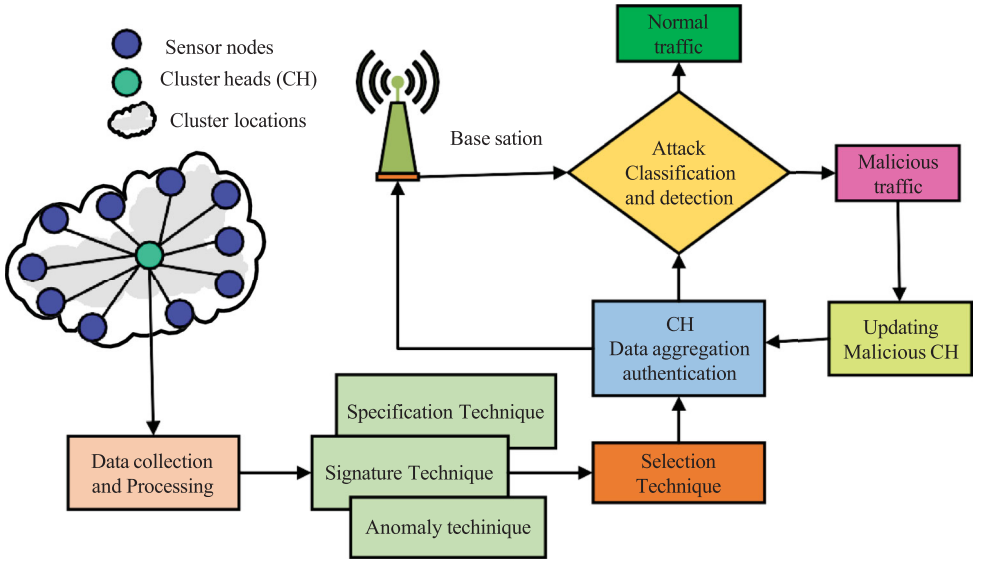


Figure 7. Hierarchical hybrid intrusion detection technique in wireless sensor networks using three phases of attack detection.

The flow-based intrusion detection scheme identifies network connections, hosts, users, and applications. Anomaly detection analyses recorded network traffic for training and detection. Cross-layer characteristics connect with packet-based IDS, extracting routing information for forwarding nodes. Hybrid IDS combines flow-based and packet-based techniques to reduce false positives and improve detection accuracy in WSNs. Correlating flow-based and layer IDS effectively detects DoS and sinkhole attacks. L. Moulad et al. (Moulad et al., 2019) proposed a hierarchical hybrid IDS technique for WSNs, using support vector machines and anomaly detection-based clustering and specification methods to detect and classify DoS attacks. The specification-based technique adopts a statistical model and behavioral/manual specifications to identify normal and malicious data, detecting anomalies and intrusion attacks in the network as shown in Figure 7. The figure illustrates the specification, signature, and anomaly detection techniques.

(Swarna Priya et al., 2020) proposed a deep neural network (DNN) model using hybrid principal component analysis and grey wolf optimisation (PCA-GWO) for intrusion detection in wireless sensor networks for medical applications. The model utilises autoencoders in a fully connected DNN with trainable parameters for training and testing. (Kanna & Santhi, 2022) examined hybrid deep learning models for effective intrusion detection systems using MapReduce based on black window-optimised long-term convolutional memory with feature selection techniques.

(Abduvaliyev et al., 2010) proposed a hybrid intrusion detection system for WSNs, utilising clustering and aggregation techniques to improve security and energy performance. Misuse and anomaly detection techniques are used to enhance detection accuracy and rate, with misuse detection targeting known attack patterns and anomaly detection employing automated training for abnormal behaviours as shown in Figure 8.

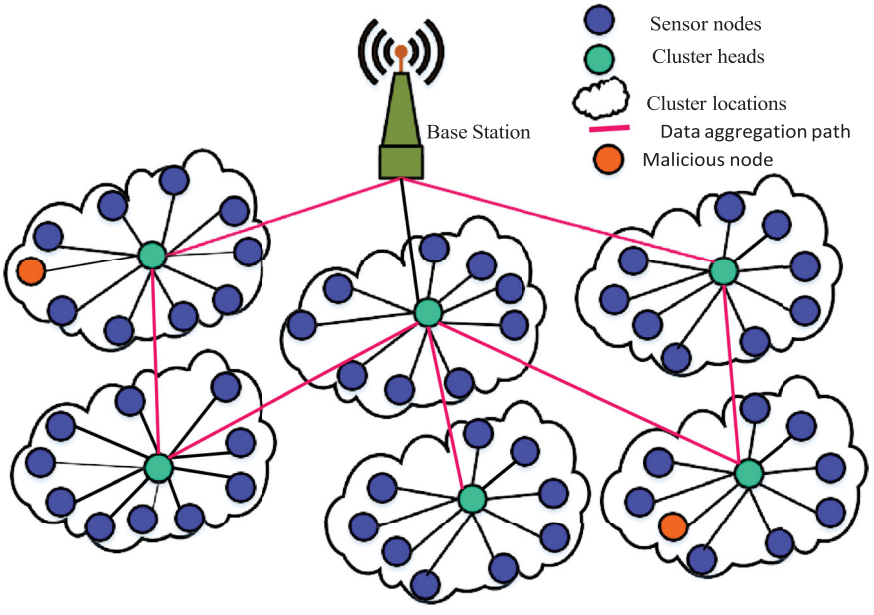


Figure 8. Illustration of data aggregation and clustering in WSNs.

Misdetection, a hybrid technique, effectively detects blackhole attacks by utilising K-medoid clustering on a synthetic data set (Ahmad et al., 2019). This approach employs the K-medoid individualised clustering method to identify anomalies caused by diversion and blackhole assaults. Additionally, a novel SDN-based Hybrid Clone Node Detection (HCND) technique has been developed for Wireless Sensor Networks (WSNs) by (Devi & Jaison, 2020). This technique proactively identifies cloned nodes using software-defined networking, ensuring the maintenance and enhancement of Quality of Service (QoS) limitations in the WSN. Hybrid multi-tiered IDS detects cyber-attacks on vehicular networks by reducing energy consumption and malicious anomaly nodes (Yang et al., 2022). Hybrid IDS detects cloning attacks in WSNs (Devi & Jaison, 2020) and classifies IoT-based security attacks for healthcare applications using feature selection and hybrid DT-GA (Saif et al., 2022).

2.2. Hybrid machine learning techniques

(Rabbani et al., 2020) proposed an effective intrusion detection system that combines machine learning with traditional security techniques. This scheme ensures mathematically secure communication among nodes and detects malicious behaviour using data pre-processing and recognition modules. Recognition techniques involve training and prediction using an optimised probabilistic neural network (PNN) with the UNSW-NB15 dataset containing normal and malicious data traffic. The hybrid PSO-PNN scheme builds a self-optimised network using particle swarm optimisation (PSO) to reduce misclassification errors and increase classification accuracy in the PNN system. PSO adapts the PNN structure, making it a self-adaptive network model using swarm behaviour patterns. Figure 9 depicts the architecture of the proposed system.

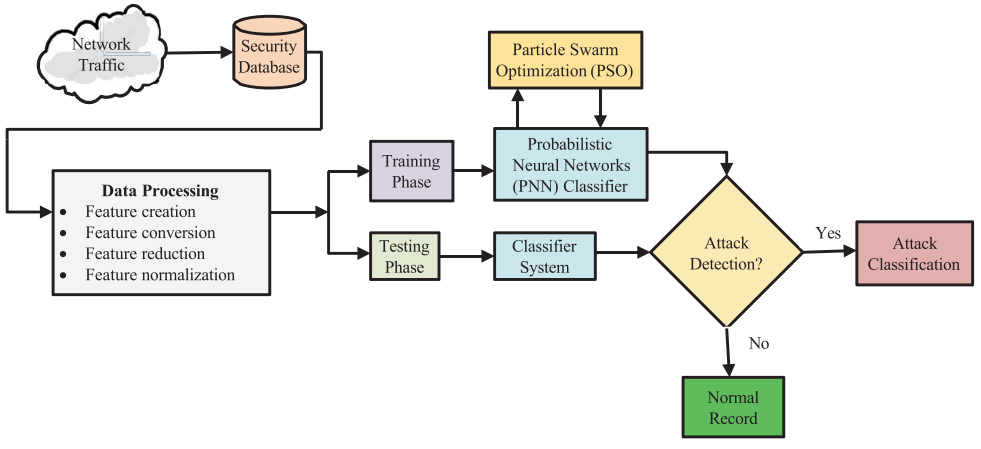


Figure 9. Hybrid PSO-PNN technique for attack classification and detection.

The features of the data network traffic are collected from the raw network packets using tools including Netmate, BRO-IDS, and Argus. The noisy features are removed to detect malicious attacks in WSNs effectively. Numeric values and symbolic variables represent the necessary features. The numeric and symbolic representations are normalised and transformed using the statistical characteristics of Equation (3).

$$Z_{normalized} = \frac{(Z - \min(Z))}{(\max(Z) - \min(Z))} \tag{3}$$

Where Z is the feature value, min (Z) is the minimum value, and max (Z) is the maximum value from the feature of the samples in the dataset.

S. M. Kumar (Kumar, 2022) presented an optimised hybrid deep neural network using a feature section algorithm for improving the intrusion detection of attacks using benchmark datasets UNSW-NB15 and NSL-KDD. The selected features are processed into the convolutional neural network consisting of layers using distance measurement and the correlational coefficient for input packets. The distance between two data points (a_i,b_i) with x and y input data, D, for arranging and selecting the features is given by Equation (4) as shown below.

$$D(x,y) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \tag{4}$$

A convolutional neural network is used to compute the selected features. The outputs are incorporated into long-term memory configuration and modified materials to improve classification precision.

(Mahajan et al., 2022) explored hybrid machine learning and deep learning techniques for network traffic analysis and classification in wireless sensor networks, including attack detection using benchmark datasets. (Faysal et al., 2022) proposed machine learning techniques to detect IoT-based WSN attacks using benchmark datasets. Hybrid eXtreme Gradient Boosting and Random Forest (XGB-RF) were effective in detecting botnet attacks with feature selection and classification using various metrics. In the same year, (Alghamdi,

2022) introduced a novel optimiser technique called PO-CFNN for IoT-based IDS. The PO-CFNN method involves three stages: preprocessing, classification, and parameter optimisation, transforming networking information into a more usable format for intrusion identification.

(Sadikin et al., 2020) presented a research study on a hybrid intrusion detection system for ZigBee-based IoT systems. They combined anomaly and rule-based machine learning techniques to detect attacks. A hybrid Long Short-Term Memory Network (LSTM) and Convolutional Neural Network (CNN) learning approach extracts network traffic features using a hybrid IDS with the CICIDS2017 dataset, achieving 99.50% overall accuracy for attack type detection (Sun et al., 2020). In another study, a hybrid classification strategy combines Kalman filter (KF) and Extreme Learning Machine (ELM) to train a predictive classifier on the sink node. It detects random WSN anomalies with promising results using normal and faulty datasets (Biswas et al., 2019).

For binary classification of attack detection, Hybrid k-means and Support Vector Machine (SVM) reduce training and testing times while maintaining high classification accuracy (Rose et al., 2020). Federated learning techniques are utilised to create a privacy-friendly framework across multiple devices using benchmark datasets (Liu et al., 2022). Hybrid optimisation and deep-learning-centric intrusion detection systems are deployed in IoT-enabled smart cities using the Hybrid Chicken Swarm Genetic Algorithm (HCSGA) method (Gupta et al., 2022). The proposed solution involves pre-processing the dataset, feature selection with HCSGA and K-means, and classification with the Deep Learning-based Hybrid Neural Network (DLHNN) classifier using the NSL-KDD benchmark dataset. Hybrid machine learning techniques utilise sampling methods and feature selection analysis to achieve better detection accuracy (Cao et al., 2022). To address sample imbalance, a hybrid sampling method combining ADASYN and RENN is employed. Hybrid deep learning methods have shown effective identification of malicious attacks when tested on performance benchmark datasets (Ullah et al., 2022).

2.3. Hybrid anomaly detection

(Umarani & Kannan, 2020) proposed hybrid anomaly detection techniques based on artificial immune systems using hybrid tissue growing techniques in wireless sensor networks to detect malicious traffic. (Yin et al., 2019) presented an anomaly detection technique that recognizes and separates normal and abnormal behaviours based on patterns of normally labelled behaviours. Data mining techniques like regression analysis, clustering analysis, outlier detection, and classification are used to extract valuable knowledge for identifying patterns of malicious nodes, improving detection accuracy and efficiency. Anomaly detection requires a machine-learning model with human effort and is error-prone. The rest of the related works are summarised as in Table 1.

Despite the existing research efforts in developing IDS based on hybrid machine learning techniques in hierarchically wireless sensor networks, some research gaps still need to be addressed. Additionally, there is a need for further exploration of the optimal combination and configuration of machine learning algorithms in the hybrid approach. Different algorithms may have varying strengths and weaknesses depending on the characteristics of the WSNs. Investigating the most effective combinations and configurations can improve detection accuracy and efficiency. These include.

Table 1. Summary of different hybrid techniques from previous works.

Ref.	Security Technique	Research Findings
Vinitha et al. (2019)	Taylor-based Cat Salp swarm Algorithm	The LEACH protocol provides an energy-efficient multi-hop routing for reliable data transfer. Throughput, energy, and delay may all be accurately measured with this method.
Singh et al. (2023)	deep learning architecture based on a fully connected feed-forward Artificial Neural Network	R = 0.78 and RMSE = 41.15 were obtained for a Gaussian sensor distribution. In contrast, R = 0.79 and RMSE = 48.36 were found for a uniform sensor distribution, indicating that the model accurately predicts the number of k-barriers for both distributions.
Singh et al. (2021)	Machine learning approach based on Gaussian Process Regression (GPR) model	An analytical method is used to extract these characteristics. Based on the simulation results, the technique provides the most accurate prediction of the k-barrier probability, with a correlation coefficient (R) of 0.85 and a Root Mean Square Error (RMSE) of 0.095.
Singh et al. (2022a)	Automated Machine Learning using Bayesian optimisation.	Compared to other explainable machine learning models, they discovered that the Gaussian process regression performed exceptionally well, with a correlation coefficient of 1, a root mean square error of 0.007, and a bias of 0.006.
Ahmad et al. (2019)	Misdetection and K-medoid clustering	K-medoid clustering with a synthetic data set proves to be an efficient way of spotting hybrid black hole attacks.
Devi and Jaison (2020)	SDN-based Hybrid clone node detection	The hybrid clone node detection detects the cloned node using proactive and verification processes based on software-defined networking in WSNs, maintaining and enhancing the quality of service.
Singh et al. (2022b)	log transformation and feature scaling on the feature set and trained the tuned Support Vector Regression (SVR)	They discovered that the model accurately predicts the number of barriers with a correlation coefficient (R) equal to 0.98, a root mean square error (RMSE) equal to 6.47, and a bias equal to 12.35.
Davahli et al. (2020)	Genetic Algorithm (GA) and grey wolf optimiser(GWO)	Reduce the dimensionality of wireless network traffic using selective features for the Internet of Things intrusion detection system with the SVM classifier.
Sun et al. (2020)	Convolutional neural network (CNN) and long short-term memory network (LSTM)	Extracts the network data traffic features using the hybrid IDS with the CICIDS2017 dataset for evaluation and achieves an overall accuracy of 99.50% for attack type.
Biswas et al. (2019)	Kalman filter (KF) with extreme learning machine (ELM)	A hybrid classification technique to train the sink node using a predictive classifier. The scheme is evaluated using the detection of random WSN anomalies data with the normal and faulty datasets.
Ren et al. (2019)	hybrid data optimisation	The practical technique uses data sampling and feature selection using a Genetic algorithm and random forest classifiers using the optimal training UNSW-NB15 dataset.

(*continued*).

Table 1. Continued.

Ref.	Security Technique	Research Findings
Regan and Leo Manickam (2019)	The optimised hybrid security model	Hybrid optimised technique for detection of malicious attacks using the hybrid secured model.
Moon and Ingole (2015)	IDS-Secured hybrid approach	The scheme provides a unique security technique for preventing and detecting attacks. The scheme realises data integrity, authentication, and energy minimisation.
Deepa and Latha (2019)	Hybrid hierarchical secure routing using clustering	The scheme selects a hybrid hierarchical secure algorithm for detection and packet delivery using a coordinate cluster head.
Sakthivel and Chandrasekaran (2018)	Hybrid security using dummy packets	A secure routing framework for detecting malicious attacks using routing protocols and dummy packet optimisation.
Yang et al. (2022)	Hybrid multi-tiered IDS using machine learning	It is practical for detecting internal and external cyber-attacks targeting vehicular networks using the CICIDS 2017 dataset.
Rose et al. (2020)	Hybrid k-means and support vector machine	Reduced training and testing times with promising classification accuracy of attack detection and classification in the network.
VenkataRao and Ananth (2021)	Hybrid optimisation and secure clustering protocol	Provides better performance using hybrid secure clustering protocol and k-means clustering for attack detection.
Gupta et al. (2022)	Deep learning-based hybrid neural network	Effective detection and classification of attacks IoT enable networks to utilise the hybrid chicken swarm genetic algorithm method.
Cao et al. (2022)	Hybrid sampling method	Provides better accuracy using feature selection analysis techniques.
Das and Namasudra (2022)	Hybrid encryption method	It is better to improve security performance in IoT-based healthcare infrastructures.
Devi and Jaison (2020)	Clone detection technique	Efficient for detection and verification of cloning attacks in WSNs.
Ullah et al. (2022)	Hybrid deep learning	Proficient in the detection of malicious attacks using benchmark datasets.
Reshma et al. (2022)	The hybrid Neighbor discovery protocol	Maximizes energy efficiency and improves security performance for detecting malicious nodes based on hybrid machine learning.
Saif et al. (2022)	Hybrid IDS approach using feature selection	Utilised hybrid DT-GA to detect and classify security attacks on IoT for healthcare applications with reduced cost.

1. **Limited evaluation on real-world WSN deployments.** Many studies have focused on simulation-based evaluations or used benchmark datasets. Future research should include more experimentation and evaluation of real-world WSN deployments to validate the effectiveness of the proposed hybrid IDS techniques.
2. **Scalability and resource constraints.** Hierarchically structured WSNs often operate under resource-constrained environments. There is a need for IDS solutions that can handle scalability issues and optimise resource usage while maintaining high detection accuracy.
3. **Dynamic adaptation.** WSNs are subject to dynamic environmental changes and evolving attack patterns. IDS solutions should be capable of dynamically adapting to these

changes and updating their detection models to ensure continuous and effective protection.

The literatures show that there is still a research gap in comprehensive studies that address the unique challenges and opportunities presented by the hierarchical structure of WSNs. Furthermore, exploring the optimal combination and configuration of machine learning algorithms in the hybrid approach is crucial for achieving improved detection accuracy and efficiency in WSNs. Future research should aim to fill these gaps and provide valuable insights into the design and implementation of advanced IDSs for hierarchically structured WSNs.

3. Network models and clustering techniques

The base station, cluster head (CH), and sensor nodes comprise the hierarchically distributed wireless sensor networks (SN) network paradigm. In this setup, the sensor nodes use a wireless connection to communicate with the cluster node and the sink node (Ghugar et al., 2019). When designing and planning the network model, the following assumptions are incorporated as shown below.

- Every mobile sensor node can roam freely within the network area (Pajila et al., 2021).
- The sensor nodes are deployed randomly.
- All of the sensor nodes are the same in every way.
- Any location within the network's range is possible for a sink to be installed.

As a result, the position of unknown nodes in the network can be calculated using beacon and sink nodes, both aware of their position and location. When it comes to routing assaults, the WSN's nodes are unprotected. In most cases, this kind of attack shortens the lifespan of the sensor nodes and causes them to run out of juice. Tunnels created by routing assaults distort the route path and use routing resources. To protect against denial-of-service and routing attacks, the proposed network model incorporates node-level security measures. The proposed system uses cutting-edge intrusion detection systems founded on hybrid machine learning approaches to identify and pinpoint cyberattacks.

The proposed network model includes five types of nodes: sensor nodes, malicious nodes, central nodes, cluster nodes, and sink nodes shown in Figure 10. The CH acts as a root node to prevent malicious communication, and the central node and CH serve as the backbone for communication with the BS. The CH uses the isolation table to conserve energy and detect attacks, with primary and secondary cluster heads for intrusion detection (Ismail & Amin, 2019). The CH also avoids depletion energy using the isolation table for attack detection. Two primary and secondary cluster heads for intrusion detection of attacks.

The attack models provide a graphical representation of the network topology, along with details about key identities and routing information that can be used to identify and exploit security flaws in the system. As seen in Figure 11, it is presumed that they have limited means and intellect to interrupt network traffic. Several variants of WSN attacks are simulated to test the effectiveness of the proposed IDS—the attack model measures how well and securely the system functions.

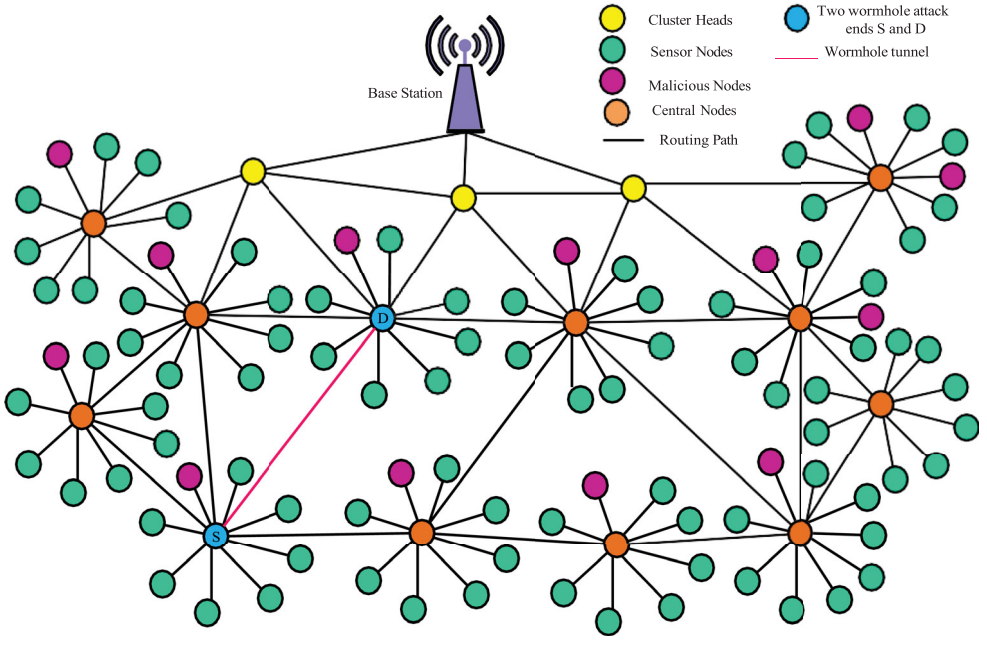


Figure 10. Hierarchical topology and configuration model for secure wireless sensor networks.

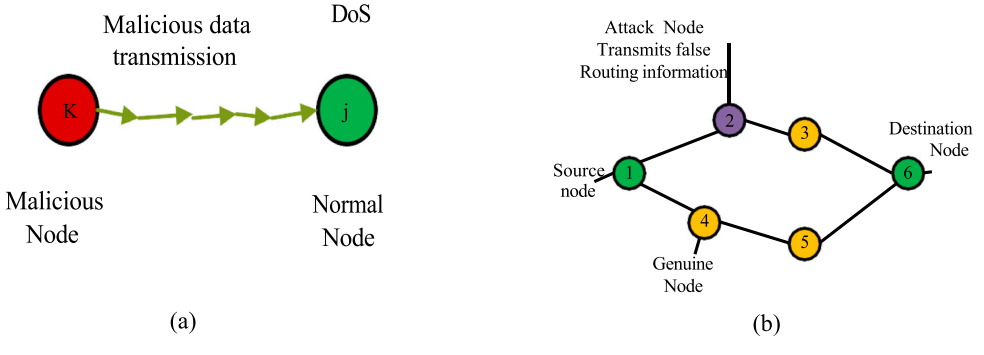


Figure 11. Jamming (a) and Sinkhole (b) attacks at the physical and network layers. (a) Jamming attacks and (b) Sink hole attacks

The application layer defines the network's threads, the Media Access Control layer, the Physical layer, the Transport layer, and the Network layer (Zou et al., 2016). Jamming security threats continuously sends harmful data, disrupting short-range connections. The transmission of jamming signals causes legitimate user and service blocking. Following is a mathematical model of an attack as in Equation (5).

$$I_a = e_i + m_i \tag{5}$$

Where I_a is the transmitted information depending on the IDS, that can be correct or incorrect, e_i is the expected information, and m is the malicious content information (Gebreyesus, 2021). The data is detected as malicious nodes have considered the network's transmitted data and energy auditing.

The channel priority is the major factor in the medium access control layer. The malicious nodes modify and change the back-off time using the manipulation approach. The attackers advertising false information in the network affects the layer routing information like the minimum hope count.

4. Methodology

The proposed system starts with system design and simulation for generating and extracting datasets. Since getting real datasets in WSNs is difficult, we use standard datasets to evaluate the effectiveness of the new advanced intrusion detection system based on hybrid machine learning models. The usefulness and efficiency of the proposed enhanced intrusion detection method based on hybrid machine learning on various classes of assaults are demonstrated by utilising publicly available datasets (Wu et al., 2020). The KDDCup 99, NSl-KDD, UNSW_NB15, and CICIDS2017 datasets are extensively used for academic study and research for attack detection evaluation as benchmark datasets. The system's effectiveness is tested using the KDD Cup 99 datasets and the intrusion detection techniques. These datasets aim to measure the IDS using a predictive decision model. The 1999 DARPA dataset is also used in this work. The dataset is evaluated using offline and real-time evaluation modes. The data is handled in several modes to establish the normal functioning of the network.

4.1. Benchmark datasets

Some of the issues with the KDD'99 data set are discussed, and a data set called NSL-KDD is proposed as a solution (Meena & Choudhary, 2017). Although it is a new and standardised genre of the KDD data set, it still suffers from some of the problems studied by McHugh. It may not be a perfect illustrative of existing real networks. Still, it is used and applied effectively as a standard data set to help researchers compare the various network-based IDSs. In addition, the NSL-KDD training and testing sets have a manageable quantity of records. The cost of doing trials on the entire dataset, as opposed to a sample, is reduced by using this method. By combining the KDD'99 Data Set with the NSL Data Set, we get several advantages over the original KDD data set.

- It avoids training classifiers on duplicate or redundant records by omitting them from the train set.
- The proposed test sets don't reuse any records, thus, the approaches with higher detection rates on common data won't unfairly boost the learners' performance.
- The proportion of records in the original KDD data set is inversely proportional to the number of records chosen from each group of challenging levels.
- Since there aren't many records in the train or test sets, we can afford to perform the experiments on the whole set without randomly picking a subset.

The NSL-KDD dataset is also used as a benchmark to test the detection performance of the proposed system using the semi-supervised machine learning (Praveen Kumar et al., 2019) models for a class of attacks with 42 attributes and class labels. Forty-one attributes are classified into content, host, traffic, and basic features. The dataset has a total record

Table 2. Frequency distribution of various attack classes in the NSL-KDD with training and testing samples for testing performance (Li et al., 2018).

samples	Normal	DoS	Probes	U2R	R2L	Total
Training samples	67,343	45,927	11,656	52	995	125,974
Testing samples	9,711	7,458	2,421	200	2,654	22,544
Total number	77,054	53,385	14,077	252	3,649	148,518

Table 3. Provides a detailed technical description of the four types of attacks in the dataset.

Attack	Attack description in the dataset
DoS	The attacker makes the network busy and denies the legitimate user access.
R2L	The intruder tries to gain access to the network or machine for a specific version of the FTP.
U2R	The attacker accesses the system's root and makes unauthorised attempts to the network.
Probe	It endeavours to assemble the data behind evading the security of the system.

of 148,515 samples sectioned into 80% of training and 20% of testing samples, as shown in Table 2, with four different classes of attacks. The vector features are extracted for training by splitting the dataset into clusters as normal and abnormal. After training, the vector features are received for classification as normal and abnormal clusters.

The dataset consists of 23 classes of attack types and is clustered into four classes of attacks, including denial of service (DoS), remote to local (R2L), user to root (U2R), and probe category. The DoS attack makes the network service busy and the authorised user inaccessible from the network (Elsaid & Albatati, 2020). The U2R attack applies vulnerabilities to the host system by sniffing the passwords of the legitimate user. The R2L injects vulnerabilities remotely into the system of the network host. The probe attack scans the network for information collecting and gathering, violating the security rule. The probe and DoS attacks have multiple links, whereas the others have single links (Pande et al., 2021). Table 3 shows the description of the four classes of attacks in the NSL-KDD benchmark dataset.

The UNSW_NB15 dataset is used as a benchmark for evaluating the effectiveness of the proposed system. This dataset has hybrid synthesised attack activities and normal network traffic data (Jatti & Kishor Sontif, 2019). The IXIA traffic generator is arranged with three virtual servers for generating the UNSW_NB15 dataset containing normal and malicious activities in the network traffic. The servers are established using public and private network traffic having IP addresses with routers. The routers are configured with a firewall that filters the traffic as normal and malicious activities. The tcpdump tool is installed on routers for capturing from the IXIA tool dispersed among the network nodes utilised as attack traffic generators with normal network traffic. The frequency distribution of the class of DoS attacks is shown in Table 4 with training and testing samples.

The method's effectiveness for identifying flooding assaults in WSNs is measured against the CIC-IDSS2017 dataset. Table 5 details some of the key elements of the training and testing dataset that can be found online at the Canadian Institute for Cyber Security Research LAB. Data about network traffic, both benign and malicious, is included in the dataset. It was manufactured to serve as a plausible in-the-background activity while gathering data. Twenty-five individuals utilising a variety of protocols were used to compile the dataset.

Table 4. Frequency distribution of attacks in the dataset.

	Training samples		Testing samples	
Attack category	weight	Percent	weight	Percent
Analysis	677	.8	2000	1.1
Backdoor	583	.7	1746	1.0
DoS	4089	5.0	12264	7.0
Exploits	11132	13.5	33393	19.0
Fuzzers	6062	7.4	18184	10.4
Generic	18871	22.9	40000	22.8
Normal	37000	44.9	56000	31.9
Reconnaissance	3496	4.2	10491	6.0
Shellcode	378	.5	1133	.6
Worms	44	.1	130	.1
Total	82333	100.0	175342	100.0

Table 5. The dataset's statistical distribution, based on a subset of its attributes.

Features	Number	Mode	Mean	Features	Number	Mode	Mean
DestinationPort	485881	4855.64	4855.64	FwdHeaderLength_A	485881	318.01	318.01
FlowDuration	485881	38856281	38856281	BwdHeaderLength	485881	350.67	350.67
TotalFwdPackets	485881	12.6	12.6	FwdPacketss	485881	15533.1	15533.1
TotalBackwardPackets	485881	14.39	14.39	BwdPacketss	485881	4910.09	4910.09
TotalLengthofFwdPackets	485881	766.63	766.63	MaxPacketLength	485881	2435.6	2435.6
FwdPacketLengthMax	485881	316.83	316.83	AveragePacketSize	485881	415.2116	415.2116
FlowBytess	485771	387017.1	387017.1	AvgFwdSegmentSize	485881	74.81795	74.81795
FlowPacketss	485771	20427.24	20427.24	FwdHeaderLength	485881	318.01	318.01
FlowIATMax	485881	31827784	31827784	SubflowFwdPackets	485881	12.6	12.6
FlowIATMin	485881	49418.34	49418.34	SubflowFwdBytes	485881	766.63	766.63
FwdIATTotal	485881	38499251	38499251	SubflowBwdPackets	485881	14.39	14.39
FwdIATMax	485881	31702017	31702017	Init_Win_bytes_forward	485881	6069.22	6069.22
BwdIATTotal	485881	19764440	19764440	Init_Win_bytes_backward	485881	1388.02	1388.02
BwdIATMax	485881	13265776	13265776	act_data_pkt_fwd	485881	8.51	8.51
BwdIATMin	485881	1321377	1321377	min_seg_size_forward	485881	26.26	26.26

Table 6. Structure of the dataset with classifications of assaults.

Attacks	DoS Slowhttptest	Normal	slowloris	DoS GoldenEye	DoS Hulk	Heartbleed
Training set	1276.8	252382.6	1504.8	6307.2	127302.4	8
Testing set	319.2	63076.4	376.2	1576.8	31825.6	2
Overall	1596	315382	1881	7884	159128	10

Table 6 shows that the dataset used to develop the predictive machine learning models contains 485881 occurrences and 31 characteristics divided into training 80% and 20% testing subsets. There are five distinct varieties of DoS attacks (Bansal & Kaur, 2018), including the widely-known Slowhttptest, Slowloris, Hulk, Heartbleed, and GoldenEye. DoS attack samples used for training and testing are shown in Table 6.

Normalisation, missing value imputation, and aggregation are all part of the data processing required to rearrange the data before the training and testing phases begin. We fill in the blanks by averaging the current values (Anbarasan et al., 2020). It is possible to convert the data into binary values as 0 and 1 by using the minimum and maximum values.

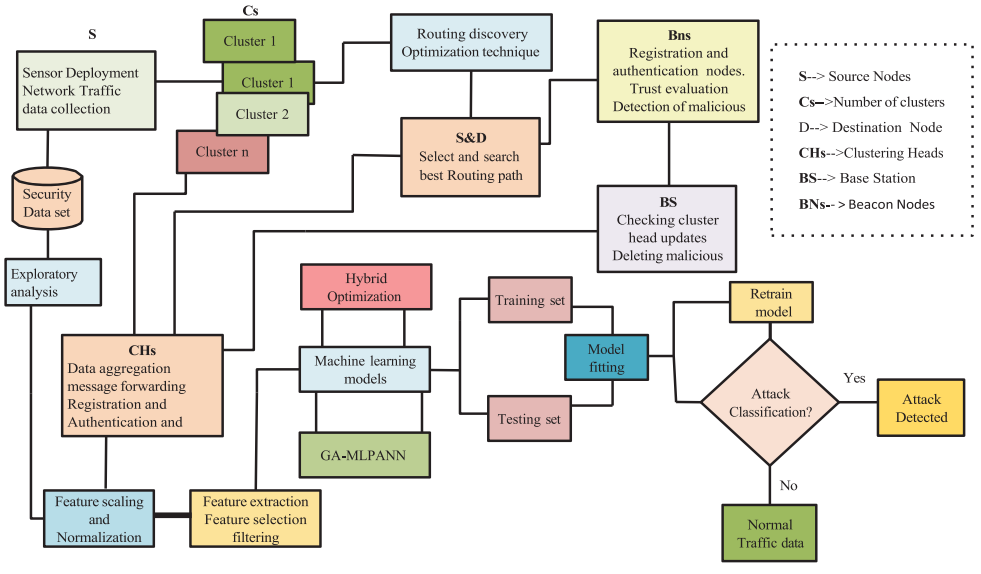


Figure 12. Framework for advanced hybrid intrusion detection system (AHIDS) Block diagram using attack detection and classification Model.

5. Proposed AHIDS framework

Figure 12 depicts how the hybrid machine learning models used in the proposed system are used to categories network data flow as either normal operation or harmful attacks. Hybrid machine-learning approaches are mostly combined to complete the intrusion detection system for attacks in WSNs (Shi & Li, 2022). Figure 12 depicts the major technical components of the proposed framework, including sensor deployment, data collection and information processing, data aggregation and clustering, decision-making, preprocessing, machine learning and optimisation for training and testing, classifier generation, and a detection and classification module for records. Since the proposed method is simple and effective, it has great potential for deployment in real-world wireless sensor networks in hierarchical clusters. Conditional control statements are a key component of the hybrid machine learning approaches for decision-making events' outcomes. The new aspect of this decision-making method is the use of a collaborative process for data analysis, which in turn aids in the automatic construction of predictive models. Decision nodes are used for prediction, while leaf nodes are used for the final classification, as seen in Figure 12, using hybrid machine learning models. Splitting training and testing benchmarks is governed by rules and reasoning generated by the hybrid machine learning models. Target categorisation is performed using the statistical metric.

Finally, the proposed system employs hybrid machine learning approaches (Gupta et al., 2022) to detect and localise attacks utilising data from both the attack and non-attack phases. Using the modified dataset, the suggested AIDS-HML learns to identify potential attacks. Assuming that all features of every sample belong to the designated class label, AIDS-HML is an efficient classification method (conditional independence assumption). Enhanced Advanced Hybrid Machine Learning (AIDS-HML) is a more advanced and hybridised version of machine learning.

Designing advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks can offer several advantages and disadvantages. Here are some potential benefits and drawbacks of the proposed system to consider.

Advantages.

1. **Improved Detection Accuracy.** Hybrid machine learning techniques combine the strengths of multiple algorithms, such as neural networks, decision trees, or support vector machines. This combination can enhance intrusion detection accuracy by leveraging each technique's unique capabilities.
2. **Adaptability to Dynamic Environments.** Hierarchically wireless sensor networks often operate in dynamic environments where network topology, traffic patterns, and intrusion characteristics can change over time. Hybrid machine-learning approaches can adapt to these changes and update their detection models accordingly, leading to better performance in dynamic scenarios.
3. **Enhanced Scalability.** Wireless sensor networks may consist of many nodes, making scalability a critical factor. Hybrid machine learning techniques can handle large-scale networks more effectively by distributing computational tasks among nodes and optimising resource utilisation.
4. **Reduced False Positives and False Negatives.** By combining multiple machine learning techniques, hybrid systems can mitigate individual algorithms' weaknesses, reducing false positives (incorrectly identifying benign activity as intrusion) and false negatives (failing to detect actual intrusions).

Disadvantages.

1. **Increased Complexity.** Implementing hybrid machine learning techniques in intrusion detection systems adds complexity to the design and deployment process. Combining different algorithms and managing their interactions requires expertise in both machine learning and wireless sensor network domains.
2. **Higher Computational Demands.** Hybrid systems may require more computational resources compared to single-method approaches. The processing and memory requirements can be significant, particularly in resource-constrained wireless sensor networks, which can impact system performance and energy efficiency.
3. **Training and Maintenance Overhead.** Hybrid machine learning models typically require more extensive training and maintenance processes. Ensuring accurate model updates, handling concept drift (changes in intrusion patterns over time), and managing retraining procedures can be time-consuming and resource-intensive.
4. **Increased Vulnerability to Attacks.** Advanced intrusion detection systems may become targets for attackers seeking to manipulate or evade detection. Hybrid machine learning models can be susceptible to adversarial attacks, where attackers exploit vulnerabilities in the learning algorithms to deceive the system. Robustness against such attacks must be considered during system design.

5.1. Sensor deployment and routing techniques

Sensor nodes are deployed based on various network models for attack detection and classification attacks in WSNs. However, WSNs face significant routing difficulties because of their restricted power supply, poor transmission bandwidth, less memory capacity, and processor capacity (Praveen Kumar et al., 2019). Due to limitations like short battery life, small memory, and low processing power, an adversary can quickly target individual nodes of WSNs when deployed in a dangerous area (Rouissi et al., 2019). It is crucial to identify malicious attacks to prevent being tricked by the adversary's fabricated data supplied by compromised nodes. Here, we distinguish between internal and external attacks on WSNs. The goal of the external attack is to reduce the effectiveness of the WSN and is carried out by parties outside of the network. Therefore, we shall elaborate on the proposal that protects against routing attacks with data while maintaining its integrity. This paper uses HML methods for WSNs to create safe protocols for extracting features and locating new routes in moderately complex hybrid and tree network topologies. The following are a few of the many advantages that machine-based routing brings to WSNs.

- Without requiring re-programming, machine learning techniques can adapt to new environments and select new CHs for routing in WSNs.
- Hybrid machine learning models can be used for various purposes in WSNs, including optimal routing, reducing communication overhead, and delay-awareness.

In this study, we employ the GA-ANN method for detecting wormhole assaults and determining energy-efficient and robust routing for WSNs. WSNs utilise GA-ANN to train their protocols based on a wide range of inputs, including residual energy, node distances, routing discovery, path selection, feature extraction, cluster heads (CH), border nodes, and the sink or base station. An enormous training set is produced, and even ANN is provided with effective threshold values for picking a set of trustworthy CH via backpropagation. Data loss in WSNs can be prevented, and energy consumption among sensor nodes is balanced with this technique.

The engineering optimisation problems used in this study find the optimal solutions under special conditions for selecting cluster head and shortest path for routing in WSNs, such as design principles, resource limitations, and safety requirements (Agushaka et al., 2022) as shown in Figure 13. Typically, metaheuristic algorithms cannot directly find the solution to constraint optimisation problems. Designing and optimising WSNs pose several challenges due to the constraints of limited energy, communication bandwidth, and processing capabilities of the sensor nodes (Ovelade & Ezugwu, 2021). However, equipped with constraint-handling techniques (CHTs), the optimisers can contend with the objective function and corresponding constraints. The purpose of optimisation is to locate the optimal answer to a problem while taking into account all relevant factors. The essence of optimisation methods lies in the gradual improvement of the generated set of solutions using a set of optimisation rules and the evaluation of those solutions using a defined objective function (Abualigah, Diabat, et al., 2021).

Unknown search space, discrete or continuous search space, non-derivative objective functions, high dimensions, and non-convexity are only a few of the characteristics of optimisation problems that prevent them from being solved in a reasonable amount of time

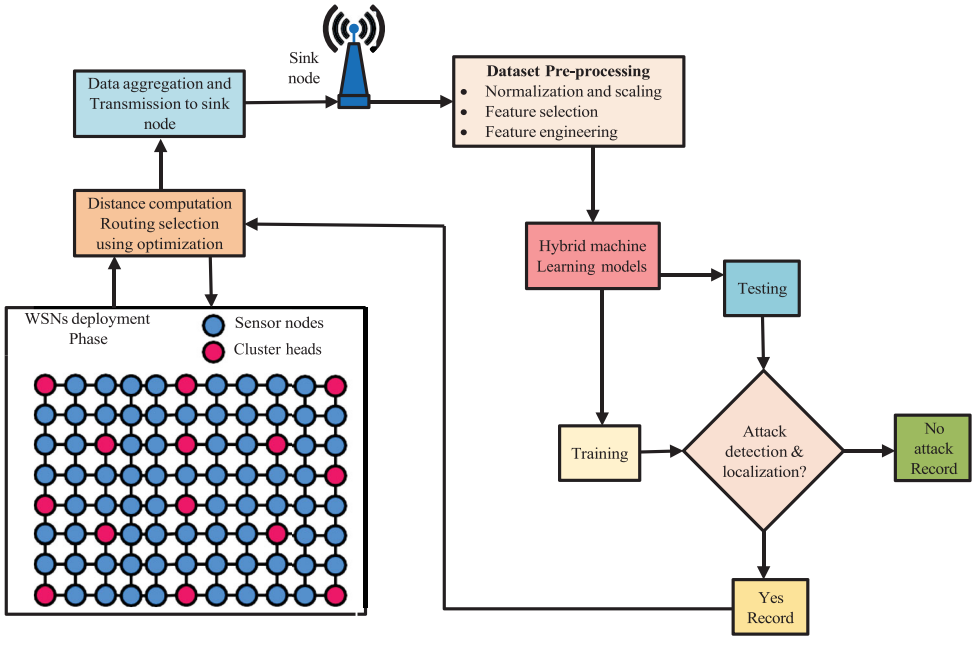


Figure 13. Illustration of flow chart diagram based on various phases of the proposed system.

using only classical methods (Ezugwu et al., 2022). These algorithms, coupled with appropriate fitness functions and problem-specific adaptations, have been used to improve the performance, reliability, and energy efficiency of IoT-WSNs (Abualigah et al., 2022). The algorithm evaluates the fitness of the candidate population using the objective function and constraints in each iteration, and the next generation of the candidate population is evaluated based on the calculated fitness function.

In the context of IoT-WSNs, an optimisation process involves finding the optimal values for specific parameters of the system in order to meet the system design requirements while minimising cost and finding the shortest path (Abualigah, Yousri, et al., 2021). The goal is to achieve an optimal configuration or solution that optimises the system's performance and efficiency. The parameters that are typically optimised in IoT-WSNs can vary depending on the specific application and design objectives. Some common parameters that are often optimised include.

- **Node Placement.** The optimal locations for sensor nodes are determined to achieve desired coverage, connectivity, and energy efficiency. This involves finding the optimal positions or coordinates for deploying the sensor nodes within the network area.
- **Routing.** The optimal routing paths are identified to transmit data from source to destination nodes efficiently. This includes finding the shortest or most energy-efficient paths considering network dynamics, congestion, and quality of services (QoS) requirements.
- **Energy Management.** Energy consumption is optimised by dynamically adjusting parameters such as sleep-wake schedules, duty cycles, or transmission power levels. The objective is to prolong the network lifetime while meeting the application requirements.

- **Resource Allocation.** The allocation of network resources, such as bandwidth and time slots, is optimised to ensure efficient utilisation. This involves determining how resources should be allocated among sensor nodes or applications to maximise overall system performance.
- **Data Aggregation.** The optimal strategies for aggregating data from multiple sensor nodes are identified to minimise redundant transmissions and conserve energy. This involves determining which nodes should aggregate data and how the data should be fused or compressed.

By employing an optimisation process in IoT-WSNs, system designers and engineers can identify the most efficient and cost-effective network configurations, enabling improved performance, energy efficiency, and overall system design.

5.2. Data pre-processing

The performance of a machine learning model is indeed influenced by the quality of the datasets on which it is trained (Singh et al., 2022b). The quality of the datasets can significantly impact the model's ability to learn patterns, generalise to new data, and make accurate predictions. Here are some key points regarding the importance of dataset quality in machine learning.

- High-quality datasets should be accurate, free from errors, and reflect the true values or labels of the target variable.
- Datasets should contain all the features and attributes required for the learning task.
- The datasets should represent the real-world problem the model aims to solve.
- Balancing the dataset or using appropriate techniques to handle class imbalance is crucial to ensure fair and accurate predictions.
- Proper preprocessing steps, such as cleaning, normalisation, and feature engineering, are essential for data preparation before training the model.
- The size of the dataset can also influence the model's performance.
- Rigorous quality assurance measures should be applied to datasets, including data validation, outlier detection, and error handling.

The data preprocessing model converts raw network traffic into the format the classification model needs in the next stage (Zhao et al., 2023). In order to offer both training and testing data, this study measures and normalises raw network traffic using a variety of data preparation approaches. A raw dataset may not have undergone any preprocessing (Roy & Chowdhury, 2021). A raw dataset is incomplete, noisy, and possibly presented unfavourably. As a result, building machine learning models from scratch using a raw dataset is impossible, as shown in Figure 14.

Preprocessing the raw data by eliminating duplicates and standardising the format improves the efficiency of a machine-learning model. This is why the training phase of a machine learning model is so important. Here are the pre-processing techniques.

Filtering: One way to clean up data is through filtering. This filter roughly estimates a desired signal pattern from a distorted signal pattern. The major goal of this filtering method is to minimise the mean square error between the estimated and intended signal patterns.

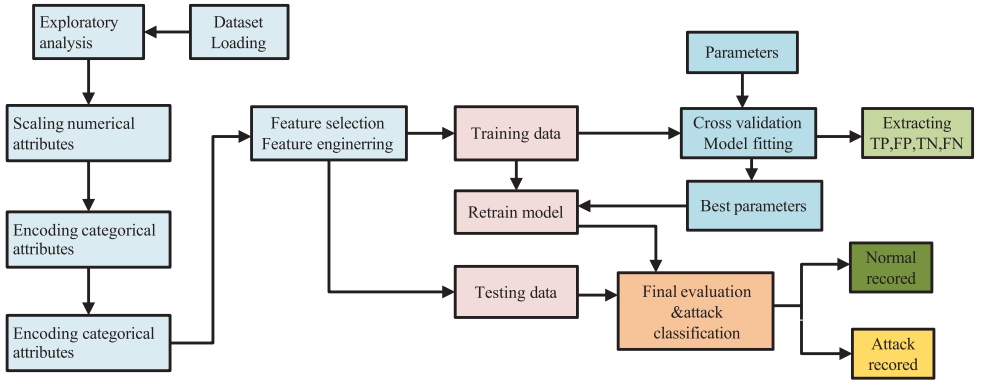


Figure 14. Data pre-processing, training and testing for model evaluation framework using benchmark datasets.

Feature Selection: Feature selection approaches are essential for choosing relevant and useful features for model learning, enhancing prediction accuracy and reducing overfitting, training time, and complexity. Common methods include Filtering, Wrapping, and embedding. Selecting the right features is crucial for data mining projects, and Figure 15 illustrates selected features using the NSL-KDD benchmark dataset (Intelligence et al., 2019). To improve application performance, identifying the best feature selection method and implementing it in relevant processes is necessary. This reduces the dataset's attributes and makes associations between them, streamlining the procedure. However, there is no universal approach to feature selection, and the dataset's condition should be considered when choosing an appropriate method. The primary challenge is finding the best feature to discriminate between classes, requiring different strategies for various datasets.

The feature selection method employs a plethora of different methods. Spearman's rank correlation coefficient formula is used for a recursive feature selection process, which then dynamically selects features, as shown in Equation (6).

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2} \tag{6}$$

Where ρ is the correlation coefficient, x_i and y_i are the feature variables and $\bar{x}$ and $\bar{y}$ are the mean values of x and y.

Feature engineering: Each pattern can be isolated with the help of a single clue provided by the feature engineering phase. When a raw dataset has a large feature set that is considered redundant, the feature extraction approach is used to create a derived set of non-redundant and informative features from the original feature set.

Windowing: Each pattern can be isolated with the help of a single clue provided by the feature engineering phase. When a raw dataset has a large feature set that is considered redundant, the feature extraction approach is used to create a derived set of non-redundant and informative features from the original feature set.

Filter Method: The filter technique employs feature ranking methods for feature selection. Ranking features indicate how crucial they are when constructing a model. The results of numerous statistical tests are used to rank the features. Each feature's connection with the

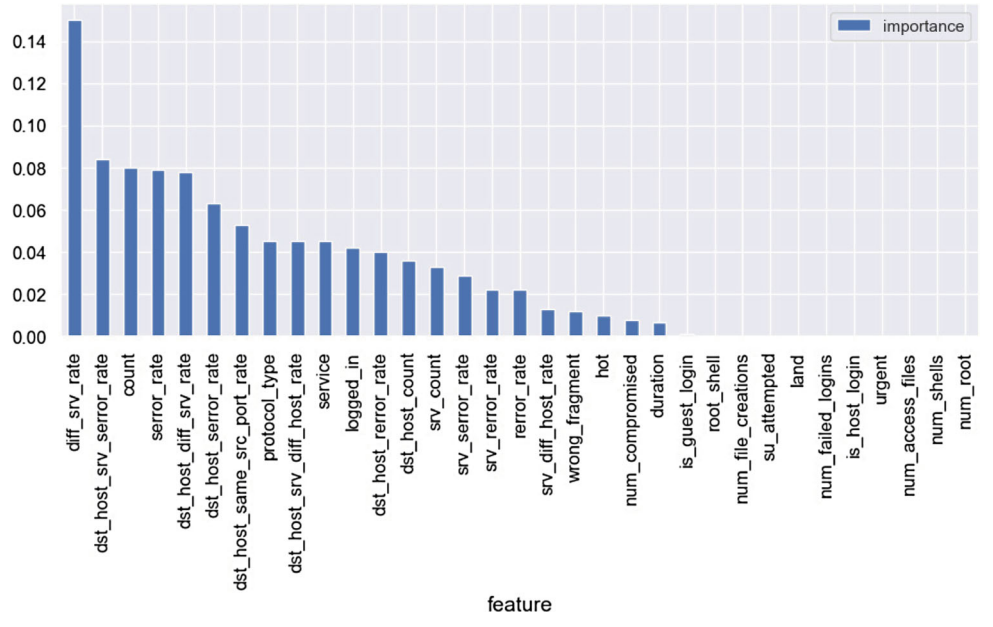


Figure 15. Selection of important feature technique using NSL-KDD benchmark dataset.

result variable is calculated in these statistical analyses. Pearson's correlation, Linear discriminant analysis, Analysis of variance, Chi-square, Variance threshold, Information gain, and so on are only some of the many statistical tests available.

Transformation and Normalization Operations: The NSL-KDD dataset utilises quantified nominal values, transforming attributes like protocol type, service, and flag numbers into numeric values (TCP = 1, UDP = 2, ICMP = 3, service = 1, flag = 2). Dataset normalisation is a crucial preprocessing technique, ensuring consistent and rescaled attribute values for effective classification (Saheed et al., 2022). Machine learning algorithms benefit from normalised data, achieving remarkable results in generalised prediction models. Minimum and maximum normalisation is used to standardise the dataset, scaling information between 0 and 1, avoiding overshadowing of lower numeric range features (Mojtaba et al., 2016). Normalisation also eliminates numerical issues in calculations, improving overall performance (Mojtaba et al., 2016; Saheed et al., 2022).

5.3. Machine learning and classification techniques

Data mining operations can utilise various data mining techniques, including hybrid machine learning methods such as Naive Bayes (NB), Artificial Neural Networks (ANN), Decision Tree (DT), Extreme Gradient Boosting (XGB), Extra Tree (ET), Random Forest (RF), Ensemble Stacking (ES), and cluster labelling K-Means (CLK-M) for attack detection and categorisation (Intelligence et al., 2019). Machine learning enables systems to learn and improve from experience without explicit programming, enhancing reliability, efficiency, and cost-effectiveness in computational procedures (Praveen Kumar et al., 2019). ML models are developed using automated and accurate processing of complex data with the extracted or chosen feature set used in conjunction with machine learning methods to

create classification algorithms. Supervised learning is suitable for standard fingerprinting data, while unsupervised or semi-supervised learning can be appropriate for crowdsourced data. Selecting the right data mining algorithm based on the dataset's structure is crucial for optimal performance. This overview outlines the use of data mining methods in classification procedures on representative benchmark datasets.

5.3.1. Naive Bayes

Naive Bayes is a basic technique for classifying data based on probability theories to identify which classes should be included. Predictions can be made after just one scan, which is straightforward. The technique is predicated on a streamlined version of the Bayesian theorem.

Conditional probability theory is used to predict to which class a given sample from a dataset will belong. Classes for test dataset samples are determined using knowledge gained during training on the training dataset. Despite its seeming lack of complexity, the Naive Bayes algorithm is highly effective. Below are the mathematical formulations for the probabilities involved in Bayes' theorem as in (7).

$$P(c|x_1,x_2,\ldots.,x_n) = \frac{P(x_1,x_2,\ldots.,x_n|c)P(c)}{P(x_1,x_2,\ldots.,x_n)} \tag{7}$$

Where P(x) is the probability of event x, c is the desired outcome, and x is the entire dataset's properties.

Based on Naive Bayes, an optimal cluster head selection method is utilised for safe and low-power routing in WSNs. An ideal collection of CHs will always maximise the network's lifetime while minimising the energy drain on individual sensor nodes. Naive Bayes ensures continued network flexibility in the face of dynamically added or modified features. This study describes a new adaptive integrated routing architecture for data collecting using a Bayesian approach.

5.3.2. Random forest

When several decision trees are trained using many different data sets, the resulting algorithm is called a random forest algorithm (RF). Breiman created this multi-technique classifier back in 2001 as an algorithm. Sub-training clusters are generated in the random forest algorithm. When forming a training cluster, preloading is used. In order to grow the trees, we employ a mechanism in which the attributes are randomly picked. The algorithm works by picking a random value from each node and utilising that as the basis for a branch as shown in Figure 16. Randomly selected factors produce the derived trees. The collected datasets are utilised as input into the Classification And Regression Trees (CART) algorithm for tree building. Each created tree is used to label the training sample and the classes assigned to which the sample is then compiled. To be processed instances are often included in the most common classification to which they belong. The RF method does not include pruning, while the CART algorithm does. An important reason why the RF algorithm outperforms the other decision tree approaches is that it doesn't rely on pruning.

The RF algorithm is quick, flexible, and more effective than alternative decision tree approaches despite using numerous tree topologies. The CART algorithm uses the GINI idex value to decide which branch to create from each node. Tree development parameters include the number of trees and the number of variables per node. The RF algorithm's

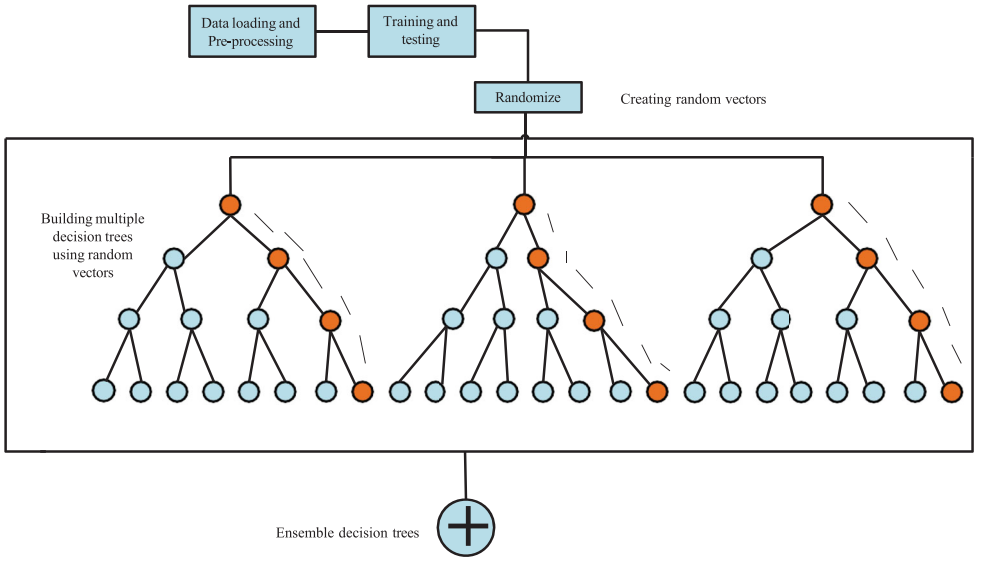


Figure 16. Block diagram of random forest operation for training and testing benchmark dataset.

basic operation is depicted in Figure 16 for attack detection and classification using training and testing the benchmark datasets.

5.3.3. Decision trees

Decision trees (DTs) are a type of supervised ML technique to classification that uses a set of if–then rules to simplify the process and improve human comprehension. The two types of nodes in a decision tree are the leaf nodes, which represent the outcomes, and the decision nodes, which represent the choices that lead to those outcomes (choice between alternatives). A decision tree can be used to predict a class or target by inferring decision rules from training data. The decision tree has the benefits of being easy to understand, helping eliminate confusion while making choices, and facilitating in-depth research. Connectivity, anomaly detection, data aggregation, and mobile sink path selection are just a few of the many problems that WSNs can address with the help of adopted decision trees.

The algorithm employs a divide-and-conquer technique. This algorithm, in contrast to ID3, incorporates normalising procedures. The algorithm determines the ratio based on the values of the information acquired. Building and repositioning intermediate trees is feasible at the time of the decision tree's inception. The decision tree method also employs branch pruning to eliminate potentially incorrect data and lower the error rate. Identifying a single node to begin the tree-building method is necessary if all of the samples belong to the same class; otherwise, the node will be labelled as a leaf and will not represent any classes. An optimal segmentation attribute is chosen if a node has characteristics from multiple classes, and the tree expands from there.

Each feature's information gain is computed, and the feature with the highest value is chosen as the tree's decision node. During the election of the cluster leader, this is the best time to identify and remove any malicious nodes. After identifying a decision node, the

procedure continues by creating a child branch off that node. If all the elements in the subgroups listed above have the same value, the procedure ends, and that value is used as the output. The process ends if the subset contains exactly one node and no distinguishing features are identified.

5.3.4. K-Means clustering

With minimal effort, the k -means method can divide a data set into a specified number of groups. Starting with a random sample of k locations, the nearest centres are assigned to each remaining point. After the data is partitioned into clusters, the centroid of each cluster is recalculated. Each time the algorithm is run, the cluster's centroid shifts until the algorithm reaches a plateau and no cluster centroid shifts. Suppose we define n as the total number of points and k as the total number of centroids. In that case, i is the total number of iterations, and d is the total number of attributes, then we can say that the time complexity of the k -means algorithm is O (n∗k∗i ∗d). In Equation (8), we see the minimisation function for the sum of squares of errors.

$$\min f(x) = \sum_{i=1}^{k} \sum_{j=1}^{N} ||x_i - y_j||^2 \tag{8}$$

Where *N* is the number of data points in the i^{th} cluster and $||x_i - y_j||$ is the Euclidean distance between x_i and y_j. The simplest clustering method, k-means, is also useful in WSNs for identifying ideal cluster heads (CHs) and detecting of malicious nodes to employ while transmitting data to the base station. This method also works well for locating productive mobile sink rendezvous spots. Choosing a different value of K can affect the outcomes in some situations. Getting the best results from the analysed data is crucial to get the value of k right. Euclidean, Manhattan, and Minkowski are just a few of the distance and neighbour node formulas that can be applied. Here are the relevant formulas given by Equation (9).

$$\begin{aligned} Euclidean &\rightarrow \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \\ Manhant\tan &\rightarrow \sum_{i=1}^{k} |x_i - y_i| \\ Minkowski &\rightarrow \left[\sum_{i=1}^{k} (|x_i - y_i|^q) \right]^{1/q} \end{aligned} \tag{9}$$

5.3.5. Hybrid-ensemble machine learning techniques

When multiple machine learning algorithms are combined into an ensemble, the resulting classification is both more accurate and faster. This approach involves several learning procedures using various machine learning approaches and then combining and categorising the results. The underlying algorithm performs two basic steps. At first, the original dataset is partitioned, and the distribution of a basic model is generated on those subsets. After doing so, the distribution is aggregated into a single model, and the results are obtained. The stacking strategy differs from standard machine learning methods because it involves a model production step. Models built from the training set are combined. You can describe the algorithm's function as follows.

- Models are created during training by employing the dataset and the training method.
- Each derived model has full annotations for all the dataset's training samples.
- The final model is built from the other models in the training dataset using the combiner method. After a final model is obtained, it categorises and tests dataset samples.
- A final prediction is made using the final model once all test dataset samples have been classified and the class predicted by the stacking algorithm of the sample is chosen.

The term ensemble technique is used to describe three distinct approaches. We're bagging, boosting, and stacking here. Data mining approaches and the capabilities of the combiner models used by each of these methods are where they diverge. Stacking strives to do both instead of maximising predictive power like boosting and minimising variance like bagging does. The function that generates a single model uses the average weight in the bagging strategy, the weighted majority vote in the boosting approach, and Logistic regression in the stacking approach.

For the suggested strategy, a tree-based Parzen estimation (PTE) is employed with hyperparameter and Bayesian optimisation (BO) techniques to further enhance the classification of the hybrid machine learning models on the benchmark dataset. Fine-tuning these parameters, or hyperparameters, is integral to every machine-learning process. Hyperparameter optimisation (HPO) enhances ML performance with decreased practitioner involvement (Feurer & Hutter, 2019). Hyperparameters are optimised in a black box and global optimisation for more accurate function evaluation. This allows us to provide a non-technical explanation of Bayesian Optimization's inner workings. Bayesian optimisation (BO) is becoming increasingly prominent in HPO for deep neural networks as a framework for the global optimisation of networks that contain expensive blackbox functions. Bayesian optimisation is a recursive method that uses a probabilistic surrogate model and an acquisition function to evaluate choices with the help of the Gaussian process. Random forest and tree Parzen estimators are just two tree-based approaches to dealing with hyperparameters (PTE). This suggested effort combines Bayesian-based optimisation (BBO) with tree Parzen estimators (TPE) to determine the optimal evaluation point for fully automated machine learning.

5.4. Performance evaluation metrics

Detection rate [18], precision, false-negative rate, and the receiver operating characteristics curve (ROC) are the performance parameters that may be measured and analysed. These indicators evaluate the system's performance, generate a categorisation report, and compare the results to those of other studies. We used a complexity matrix as one of our criteria for rating submissions (Intelligence et al., 2019). Some of the metrics used in this paper are bulleted as follow:

Routing Attack Detection Rate: This metric indicates the percentage of detected routing attacks, including blackhole, wormhole, Sybil and misdirection assaults, out of the total number of simulated attacks. A higher detection rate signifies the effectiveness of the hybrid technique in identifying malicious routing behaviours.

False Positive Rate: This metric measures the proportion of legitimate network activities falsely identified as routing attacks. A low false positive rate is desirable to avoid unnecessary alarms and ensure the reliability of the detection system.

False Negative Rate: This metric represents the percentage of actual routing attacks that were not detected by the hybrid technique. A low false negative rate indicates that the method can effectively capture most routing attacks without missing significant ones.
Precision: Precision is the ratio of true positive detections to the sum of true positive and false positive detections. A higher precision indicates that the reported routing attacks are more likely to be genuine, reducing the chances of false alarms.
Recall (Sensitivity): Recall measures the proportion of true positive detections to the sum of true positive and false negative detections. A higher recall signifies that the hybrid technique successfully identifies a larger portion of routing attacks, ensuring better coverage.
F1 Score: The F1 score is the harmonic mean of precision and recall. It provides a balanced assessment of the algorithm's accuracy in detecting routing attacks. A higher F1 score indicates a better trade-off between precision and recall.
Execution Time: This metric measures the time taken by the hybrid technique to analyse network data and detect routing attacks. A shorter execution time is preferred to enable real-time or near real-time response to potential threats.
Resource Utilization: This metric evaluates the amount of computational resources, memory, or network bandwidth required by the hybrid technique to perform routing attack detection. Efficient resource utilisation including time and energy is essential for practical deployment.

Values from the complexity matrix are used to determine the criterion for evaluation. Following is a breakdown of the values in the complexity matrix.

- In the dataset, TP (true-positive) refers to the number of samples that were accurately predicted to be incursions.
- Several samples in the normal class were correctly predicted to be in the normal class (true-negative or TN).
- False-negative (FN). The fraction of intrusions samples that were wrongly classified as normal.
- Number of normal samples in the dataset that were wrongly classified as incursions (FP, or false positive).

The detection rate is calculated by dividing the TP value by the total number of samples for which intrusion estimates were calculated. The accuracy value measures how well a system performs in classifying data by comparing the fraction of data points that were correctly labelled by the system to the total number of data points. To demonstrate the system's efficiency, we employ the following mathematical Equation (10).

$$\text{Detection Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{False alarm rate} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Accuracy(Acc)} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}$$

$$\text{F1 - score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{10}$$

It has been noted that the designed IDS can perform four different outcomes for each traffic operation. The following scenarios are generated using the confusion matrix. First, a True Positive (TP) occurs when an intrusion detection system (IDS) reports a successful detection of malicious activity on a network (Mohd et al., 2020); second, a True Negative (TN) occurs when an IDS does not report a successful detection of malicious activity, third, a False Positive (FP) occurs when an IDS reports no malicious activity, and fourth, a False Negative (FN) occurs when an IDS reports a successful detection.

Mean squared error (MSE). Mean squared error (MSE) measures the amount of error in statistical machine learning models for computing the position and distance of the wormhole attack between two points as in Equation (11). It assesses the average squared difference between the observed and predicted values of each sensor node's position and location, having its unique identity to detect routing attacks. When a model has no error, the MSE equals zero. As model error increases, its value increases. The mean squared error is also known as the mean squared deviation (MSD).

$$\text{MSE} = \frac{\sum_{i=0}^{n} (x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2}{n} \tag{11}$$

Where.

- x_i, and y_i is the i^{th} observed values.
- $\bar{x}_i$, and $\bar{y}_i$ are the corresponding predicted values.
- n is the number of observations.

The mean squared error uses a formula that is quite close to the variances. The MSE is calculated by square root, the difference between the observed and anticipated values. That should be done for every observation. After that, divide the total by the total number of observations to get the square root.

5.5. Simulation and environmental setup

Network design and model simulations were executed in MATLAB R2021a on a Windows 10 64-bit x64-based processor running an Intel Xeon Silver 4214 CPU at 2.20 GHz 2.19 GHz (2 processors), with 128GB (128GB useable) of installed RAM. Data processing and analysis with machine learning classifiers are performed in Python libraries, including Keras numpy, Sklearn, Seaborn, and pandas using Anaconda navigator and MATLAB R2021a. The simulation parameters for running network attack scenarios are depicted in Table 7, along with the values. This study assumes that Node-0 is the final destination for network traffic. A total of 5 s are allotted for the simulation.

Simulations of wormhole routing attacks are undertaken, with the results being created using an artificial neural network and a genetic algorithm that have been genetically

Table 7. WSN configuration of simulation setting.

Parameter	Setting	Parameter	Setting
Base Station	1	Topology	Hierarchical
Filed size	1500 m × 1500 m	Number of attacks	2
Number of Nodes	200	Mobility Model	Random
Protocol type	Routing	Number layers	10
Cluster size	10	Max epochs	200
Attack type	wormhole	Data size	5000 Kb
Number iterations	200	Simulation Time	5 s

enhanced for optimal effectiveness. After that, a malicious node is added to the network in order to generate and extract features for both benign and malicious network traffic. This procedure is then repeated for another five seconds of simulation time in order to develop a new database. The simulation scenarios that make use of a mobility and routing protocol that is based on a random selection of intermediate nodes and mobility nodes to identify and extract features from probable routes are outlined in Table 7. Between two malicious nodes, a wormhole attack is injected, which results in the creation of a tunnel.

6. Experimental results and analysis

In this section, each sensor node in the network is capable of establishing connections with all other nodes, resulting in a fully connected network topology. This dynamic connectivity allows the formation of wormhole tunnels. Wormhole tunnels refer to virtual tunnels or channels created between two malicious nodes in the network. as in Figure 17 20 (a) and (b). These tunnels bypass normal routing mechanisms and enable attackers to disrupt the routing discovery process. By exploiting these tunnels, attackers can carry out various routing attacks, such as selective forwarding, blackhole attacks, or Sybil attacks. The purpose of this study is to investigate the detection and prevention of such routing attacks in wireless sensor networks. By simulating the fully connected network topology and incorporating wormhole tunnels, the researchers aim to develop effective mechanisms to discover and mitigate routing attacks.

The routing wormhole attack highly affects the sensor nodes' energy consumption and timing operation for effective communication, as shown in Figure 17 (c) and (d).

The simulation results show that the proposed attack detection and classification techniques are effective, with an average detection accuracy of 99.46%, varying the hope count and wormhole tunnel of the routing attacks across the network. The results also show that hybrid techniques improve the prediction error and maximise the performance, as shown in Figure 18 (a) and (b) histograms. Figures show the differences between targets and actual outputs for computing the errors of the unknown nodes. Targets represent expected outputs, and outputs represent the actual outputs (Constraints, 2016). The error of the training data is almost 0, whereas the error of testing data is higher than that of training errors. This confirms and validates the proposed technique is effective for detecting the wormhole attack in WSNs. The validation and efficiency of the proposed system are depicted in Figure 18 (c) and (d) for detection of routing attacks at epochs 7 and 200 epochs with minimum mean squared error (MSE) of 0.0067 and $2.143x10e^{-08}$.

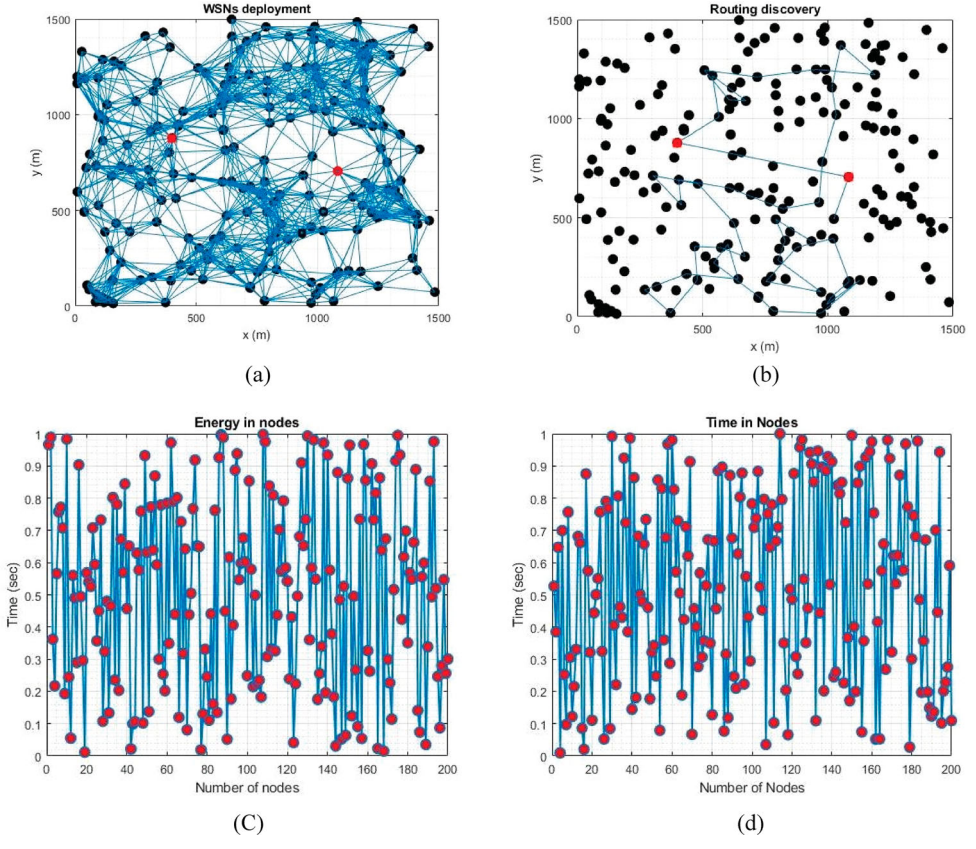


Figure 17. wireless sensor network deployment and routing discovery for analysing energy consumption and time elapsed. (a) Dynamic network deployment. (b) Routing discovery and feature extraction. (C) Energy consumed for nodes with time and (d) Time elapsed for each node.

6.1. Attack detection analysis

The samples from the reference datasets have been put through training and testing processes (Panigrahi et al., 2022). First, we randomly assign each sample to two groups. the training and test sets. Step two involves using the whole training set for both training and testing. Finally, cross-validation was utilised to test how well the proposed model actually worked. The area under the curve, false alarm rate, precision, and classification accuracy are used to evaluate performance. Machine learning models are used to assess how well the proposed method performs on benchmark datasets that simulate a variety of assaults against wireless sensor networks. When assessing the efficacy of the proposed system for detecting routing attacks in WSNs, the hybrid optimised machine learning also uses the same benchmark dataset. Table 8 provides a comparison of the results obtained by using various machine learning algorithms. Cluster labelling (CL) k-means binary classification methods are used to boost the suggested system's performance further. Table 8 shows the comparative performance of the various hybrid machine-learning techniques.

The hyperparameter and Bayesian optimisation (BO) techniques and the tree-based Parzen estimation (BO-PTE) are used to boost the performance of hybrid machine learning

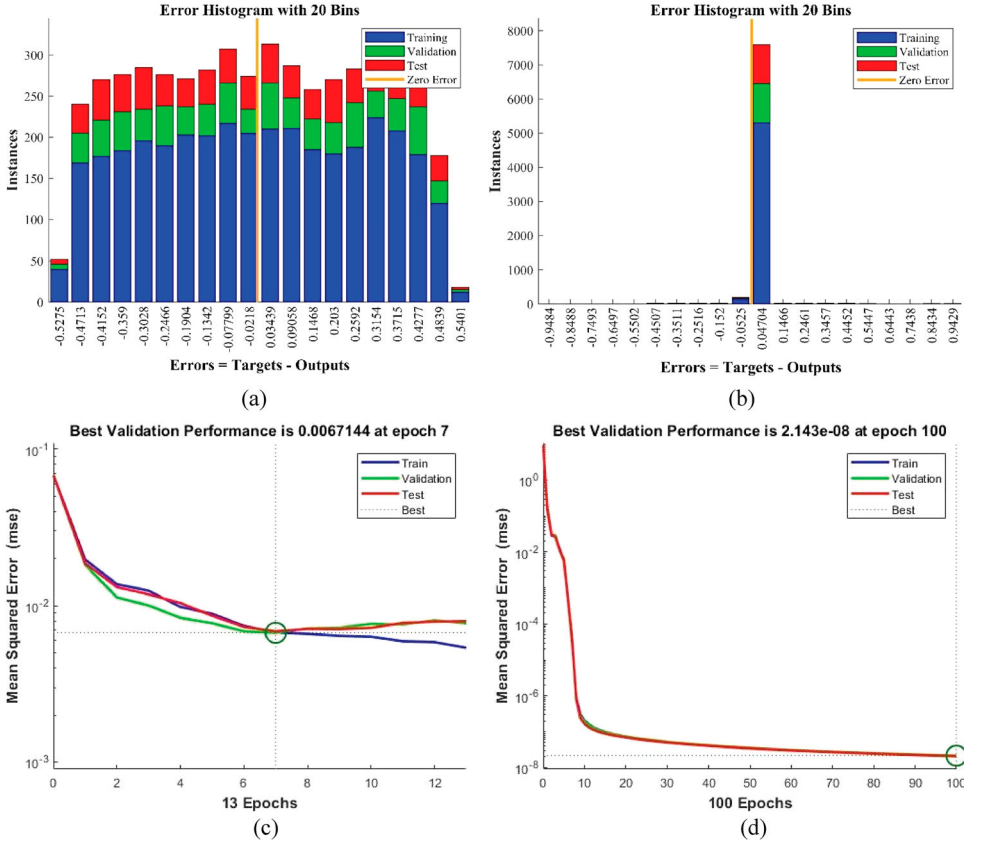


Figure 18. The proposed system's performance evaluation using hybrid GA-ANN takes 100 samples and varying the number of epochs. (a) Histogram with 300 instances. (b) Histogram with 300 instances. (c) Best performance ate epoch of 7 and (d) Best performance at the epoch of 100.

Table 8. Evaluation of hybrid machine learning models on standard datasets.

	Results against the UNSW_NB15 dataset				Results against the CICIDS2017 dataset			
Classifier	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
XGBoost	99.78	99.74	99.78	99.75	99.82	99.86	99.82	99.83
Random Forest	99.75	99.67	99.75	99.70	99.82	99.82	99.82	99.80
Decision Tree	99.68	99.63	99.68	99.66	99.82	99.91	99.82	99.85
Extra Tree	99.72	99.66	99.72	99.68	99.82	99.80	99.82	99.80
Ensemble Stacking	99.78	99.74	99.78	99.75	99.82	99.91	99.82	99.85
CLK-Means	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

models for the proposed system. Table 8 displays the results of evaluating the proposed scheme's performance using the UNSW NB15 benchmark dataset and many other machine-learning models. When applied to the benchmark dataset, the binary classification method employing hybrid cluster labelling K-means obtains a classification accuracy of 100%.

Table 9 shows how merging different hybrid machine-learning models and moving data frames from one machine-learning to the other improves the proposed system's performance even further. The results demonstrate that hybrid ML models outperform their

Table 9. Comparison of various hybrid ML models using NSL-KDD dataset for attack detection and classification.

	Performance evaluation metrics					
ML Classifiers	validation	Accuracy	Precision	Recall	F1-score	Training time
Naive Baye (NB)	83.94	83.71	90.35	83.71	85.68	0.025
Decision Tree (DT)	87.94	88.10	88.34	88.10	88.07	0.22
XG Boost (XGB)	99.16	99.34	99.32	99.34	99.32	1.83
Random Forest (RF)	98.66	99.46	99.44	99.46	99.45	0.17
DT-XGBoost	99.57	99.80	99.80	99.80	99.80	1.24
RF-XGBoost	99.57	99.80	99.80	99.80	99.80	1.44
RF-DT	99.57	99.79	99.79	99.79	99.79	0.327

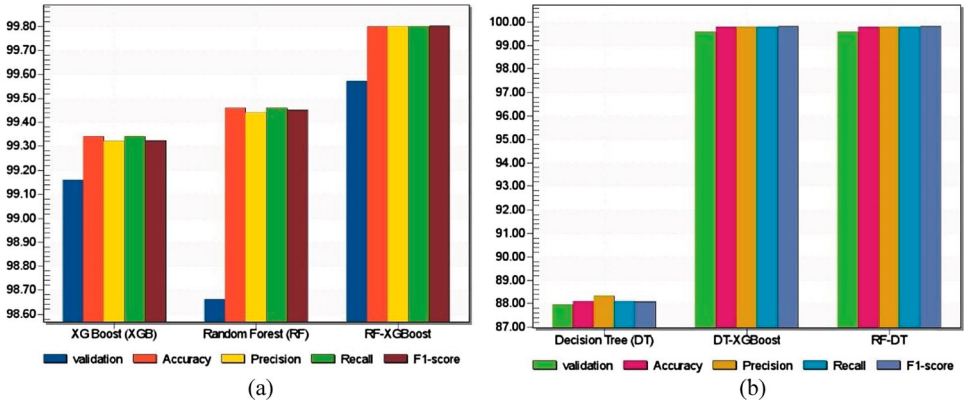


Figure 19. Performance comparison of various machine learning models using NSL-KDD. (a) RF-based comparative analysis and (b) DT-based comparative analysis

standalone counterparts in terms of Validation, Accuracy, Precision, Recall, F1 score and training time. When it comes to classification and detection, based on the results and model validation, we can say that the created system has high classification and detection accuracy against DoS attacks in WSNs.

The results show that hybrid machine techniques perform better attack detection and classification of attacks using the NSL-KDD benchmark dataset, as shown in Figure 19 (a) and (b). For attack detection and classification, the extreme gradient boosting (XGB)-an enhanced hybrid of a random forest and a decision tree achieves better results than either the random forest or the decision tree alone in terms of validity, accuracy, precision, recall, and f1-score.

The performance of the proposed technique is effective compared to L. Yang et al. (Yang et al., 2022) developed multi-tiered hybrid intrusion detection systems (MTH-IDS) for secure vehicular networks using the benchmark dataset CICIDS2017 for known and unknown attacks. They achieved average detection accuracy of 99.88% using binary classification. P. Sun et al. (Sun et al., 2020) developed a hybrid deep learning-based intrusion detection system (DL-IDS) using a convolutional neural network and a long short-term memory network (CNN-LSTM). The scheme achieved an average detection accuracy of 98.67% by extracting the network traffic. This proves that the proposed scheme effectively detects DoS attacks using the benchmark dataset in wireless sensor networks, as shown in Figure 20 (b), using

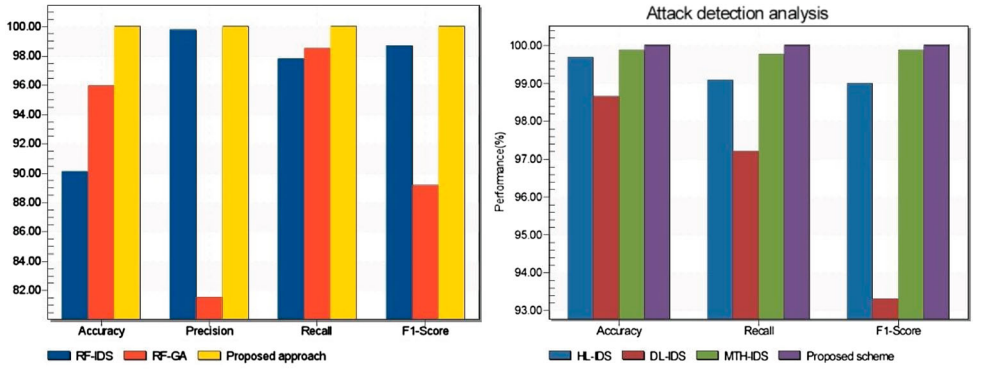


Figure 20. Performance comparison of the proposed scheme-based hybrid machine learning techniques using benchmark datasets. (a) Comparison based on UNSW_NB15 and (b) Comparison-based CICIDS2017

attack detection performance metrics. (Kasongo, 2021) presented an intrusion detection system for the Internet of Things using random forest based on a genetic algorithm (RF-GA) for feature selection, as shown in Figure 20. This achieved average detection accuracy of 87.61%, which is less than compared to 100% using hybrid binary classification. (Suleiman & Issac, 2018) Evaluated six machine learning classifiers using UNSW_NB15, phishing and NSL-KD benchmark datasets for intrusion detection system. Random forest based intrusion detection system (RF-IDS) produced better detection accuracy using UNSW_NB15. Temporal and spatial features to enhance attack detection and classification. This confirms the proposed technique is effective for detection and localisation of attacks as shown in Figure 20 (a).

In order to achieve high-performance intrusion detection across a wide range of attack types, B. Media et al. (Intelligence et al., 2019) proposed a hybrid-layered IDS (HL-IDS) that employs several distinct machine learning and feature selection approaches as shown in Figure 20 (b). The size of the NSL-KDD dataset is decreased in the created system by first performing data preprocessing on the dataset using various feature selection algorithms.

G. H. Lai (Lai 2016) Proposed detecting wormhole attacks in WSNs using low-power routing protocol and achieving 100% accuracy with fixed range and wormhole tunnel points. This confirms the proposed technique is effective for localising and detecting routing attacks in wireless sensor networks using a benchmark dataset. Y. Yuan et al. (Yuan et al., 2018) Presented a novel lightweight method for Sybil attack detection in distributed WSNs using the approximate point in a triangle (APIT) localisation approach. They achieved an average detection rate of 90%, which is less than the proposed work. D. Upadhyay et al. (Upadhyay et al., 2021) proposed a framework for intrusion detection systems in smart grids using Gradient boosting feature selection by applying machine learning classification techniques. The scheme combines feature engineering with machine learning classifiers and achieves the performance as in Figure 21 (a). This confirms that the proposed method is effective for various applications of DoS attacks in wireless sensor networks.

A unique feature selection algorithm, the dynamic recursive feature selection algorithm, was introduced by Nancy P et al. (Nancy et al., 2020), which chooses an optimal number of

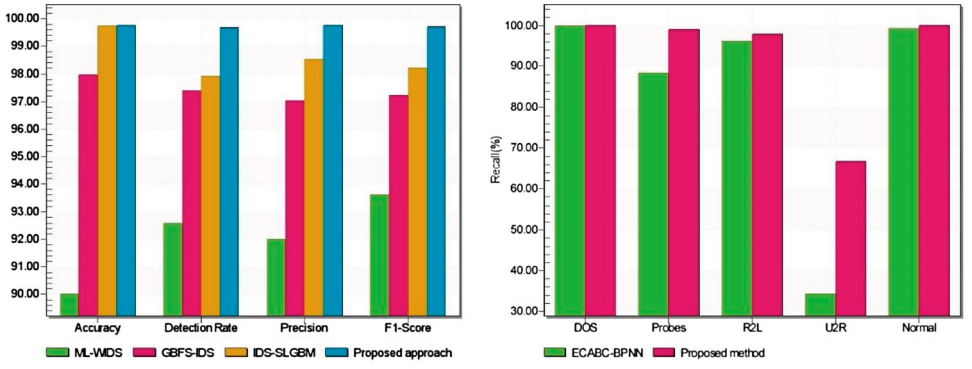


Figure 21. Performance comparison of the proposed technique with other previous works. (a) Comparative analysis of the proposed scheme and (b) Comparative analysis based on recall

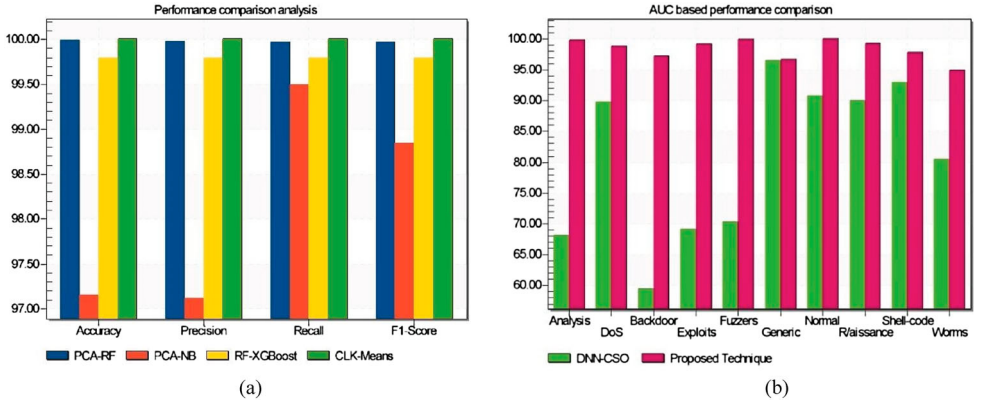


Figure 22. Examining the proposed system's performance of similar works using standard benchmarks. (a) Comparison of the proposed system with recent works and (b) AUC-based performance comparison using the UNSW-NB15 benchmark dataset.

features from the data set. Moreover, a sophisticated intrusion-detection system based on a fuzzy logic algorithm (IF-IDS) using the NSL-KDD dataset is shown in Figure 20. Extending the decision tree approach and including convolution neural networks are also presented as means by which to detect the invaders efficiently. The technique of intelligent feature selection algorithm named dynamic recursive feature selection algorithm (DRFSA) has been proposed in this work, which picks the important features to construct the data set. G. Qi, J. Zhou et al. (Qi et al., 2021) presented a new ECABC-BPNN, a combination of back propagation neural networks (BPNNs) and elite clone artificial bee colonies (ECABCs), that improves upon the standard BPNN's weight and threshold settings as depicted in Figure 21 (b).

The proposed system is evaluated using benchmark datasets based on accuracy, precision, recall, F1-score and AUC criteria. The proposed scheme's performance is effective compared to R. Khilar et al. (Khilar et al., 2022) and (Saheed et al., 2022) in terms of various evaluation metrics. This proves the system effectively detects and localises DoS assaults in WSNs, as in Figure 22 (a) and (b).

Table 10. Performance evaluation of the proposed system using the NSL-KDD benchmark dataset.

	Proposed AIDS-HML approach			Intelligent fuzzy-based IDS			ECABC-BPNN
Class of attacks	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision
DOS	99.89	99.93	99.91	99.99	97.23	98.72	98.56
Probes	99.30	98.87	99.08	92.67	97.97	93.34	86.70
R2L	98.91	97.84	98.38	57.39	28.32	42.97	97.20
U2R	80.00	66.66	72.72	95.23	63.56	59.03	83.67

The next step is using ECABC-BPNN to identify threats in a computer system's network. The comparison and Conducted experiments on assault classification using benchmark dataset as shown in Table 10.

The proposed scheme is further compared for validation with previous works to detect and localise routing attacks in WSNs. S. Jiang et al. (Jiang et al., 2020) Proposed an intrusion detection system based on a secure light gradient boosting machine (IDS-SLGBM) in wireless sensor networks using the WSN-DS benchmark dataset with the class of routing attacks.

The experimental results and analysis show that designing advanced intrusion detection systems (IDS) based on hybrid machine learning techniques in hierarchically wireless sensor networks introduces several novel aspects and contributions. Here are some key points that highlight the novelty of this design:

Hierarchical Wireless Sensor Networks: Hierarchical architecture in wireless sensor networks introduces an additional layer of complexity and organisation. The network is divided into multiple levels or tiers, with different roles assigned to nodes at each level. This hierarchical structure helps in efficient data aggregation, routing, and management in large-scale sensor networks.

Hybrid Machine Learning Techniques: The IDS design incorporates hybrid machine learning techniques, which combine multiple algorithms or approaches to enhance the accuracy and effectiveness of intrusion detection. This hybridisation can involve integrating different machine learning models, such as combining supervised and unsupervised learning methods or combining traditional rule-based techniques with machine learning algorithms.

Advanced Intrusion Detection: The IDS focuses on advanced intrusion detection, aiming to detect sophisticated attacks beyond simple rule-based or signature-based detection methods. Advanced attacks often exhibit complex patterns or behaviours that can be challenging to identify using traditional approaches. By leveraging machine learning techniques, the IDS can learn and adapt to evolving attack patterns, enabling the detection of novel and unknown attacks.

Novelty in Feature Selection: The design may introduce novel approaches to feature selection, which involves identifying the most relevant and discriminative features from the sensor data to train the machine learning models. Effective feature selection plays a crucial role in improving the accuracy and efficiency of intrusion detection systems, especially in resource-constrained wireless sensor networks.

Scalability and Efficiency Considerations: The design considers wireless sensor networks' scalability and efficiency requirements. Hierarchical structures and optimised machine-learning techniques are employed to reduce the network's computational overhead,

energy consumption, and communication overhead. These considerations ensure that the IDS is suitable for deployment in resource-constrained environments.

6.2. Limitation of the proposed system

While design of advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks presents promising approaches to intrusion detection in wireless sensor networks, it also has certain limitations that should be acknowledged:

Scalability: The performance of the proposed advanced intrusion detection system may degrade with the increase in the size of the wireless sensor network. Handling a large number of nodes and data traffic could pose challenges in terms of computational resources and memory requirements.

Complexity and Overhead: The hybrid machine learning techniques used in the system may introduce additional complexity and computational overhead, particularly for resource-constrained sensor nodes. This could impact the real-time responsiveness and energy efficiency of the overall network.

Training Data Collection: Obtaining labelled training data for machine learning algorithms in wireless sensor networks can be challenging. Collecting a diverse and representative dataset of intrusion scenarios, including rare attacks, might be difficult due to the limited resources and controlled environment.

Intrusion Diversity: As new intrusion techniques and attack patterns emerge, the detection system may face challenges in generalising and adapting to previously unseen or zero-day attacks, especially when using pre-trained machine learning models.

Security and Privacy: Deploying an intrusion detection system in the network itself could potentially become a target for attacks. Adversaries might try to manipulate the system's behaviour or exploit its vulnerabilities to evade detection.

Adaptability to Network Changes: As the wireless sensor network topology changes due to node failures, additions, or mobility, the intrusion detection system should be able to adapt and maintain its effectiveness.

False Alarms: Hybrid machine learning techniques might lead to false alarms in certain situations, triggering unnecessary responses and consuming valuable resources for investigating non-existent attacks.

7. Conclusion and future work

The proposed advanced intrusion detection system based on machine learning effectively detects and classifies attacks for scalable and manageable in hierarchically distributed wireless sensor networks. This research aims to create a classification model for an advanced intrusion detection system based on hybrid machine learning, specifically tailored for use in wireless sensor networks to detect intrusions. Each sensor node collects information on the state of its features and reports it to the cluster's central processing node. The cluster leader checks the data and then forwards it to the main cluster head. The proposed hybrid machine learning models use training and testing data to identify attacks. Our suggested IDS-HML outperforms state-of-the-art systems regarding detection and localisation

accuracy in a simulated attack on a WSN. By comparing the hypothetical outcomes to earlier research, we find that they are credible. The simulation results show that the proposed system is effective for detecting routing attacks with a localisation accuracy of 99.46% of the wormhole routing attacks. The effectiveness of the suggested system has been measured in accuracy, precision, TP Rate, FP Rate, F-Measure, Mean squared error, and Time. The designed IDS-HML achieved 99.82%, 99.91%, 99.85%, 99.82%, and 100% for average detection accuracy, precision, F1-score, recall, and CLK-Means respectively, in the presence of normal and intrusion traffic using CICIDS2017 dataset as a benchmark for multiclass and binary classifications. This work is implemented using MATLAB for network planning and simulation of attack scenarios. The Python libraries are utilised for hybrid machine-learning classification techniques. This model uses logic rules for decision-making and interpretable predictive models.

Overall, the novelty lies in the combination of hierarchical wireless sensor networks, hybrid machine learning techniques, advanced intrusion detection capabilities, novel feature selection approaches, and considerations for scalability and efficiency. These elements contribute to the development of a robust and effective IDS for wireless sensor networks.

Although the proposed method performs well, it is essential to note that IoT-based WSNs are still susceptible to attacks not addressed in this study. Since the dynamics of the attacks change with time, the topology and design should cope with the attack scenarios. The countermeasures module is only provided in concept, which is another shortcoming. Therefore, we plan to investigate and eventually offer specialised advanced hybrid intrusion detection systems for each type of assault utilising benchmark datasets to evaluate hybrid machine learning techniques. In future work, we will explore collaborative advanced intrusion detection systems based on machine learning in IoT-based wireless sensor networks for different applications using benchmark datasets for evaluations.

Declarations

Funding There is no research funding for this work.

Conflict of interest The authors declare that no competing interests or personal relationships could affect this work.

Data availability The corresponding author can provide access to the data utilised to support this study's findings upon request.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abdulganiyu, O. H., Ait Tchakoucht, T., & Saheed, Y. K. (2023). A systematic literature review for network intrusion detection system (IDS). *International Journal of Information Security*, https://doi.org/10.1007/s10207-023-00682-2

Abduvaliyev, A., Lee, S., & Lee, Y. K. (2010). Energy efficient hybrid intrusion detection system for wireless sensor networks. *ICEIE 2010 - 2010 International Conference on Electronics and Information Engineering, Proceedings*, *2*(Iceie), 25–29. https://doi.org/10.1109/ICEIE.2010.5559708

Abualigah, L., Diabat, A., Mirjalili, S., Elaziz, M. A., & Gandomi, A. H. (2021). The arithmetic

optimization algorithm. *Computer Methods in Applied Mechanics and Engineering*, *376*, 113609. https://doi.org/10.1016/j.cma.2020.113609

Abualigah, L., Elaziz, M. A., Sumari, P., Geem, Z. W., & Gandomi, A. H. (2022). Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer. *Expert Systems with Applications*, *191*(November 2021), 116158. https://doi.org/10.1016/j.eswa.2021.116158

Abualigah, L., Yousri, D., Elaziz, M. A., Ewees, A. A., Al-qaness, M. A. A., & Gandomi, A. H. (2021). Aquila optimizer: A novel meta-heuristic optimization algorithm. *Computers and Industrial Engineering*, *157*(October 2020), 107250. https://doi.org/10.1016/j.cie.2021.107250

Agushaka, J. O., Ezugwu, A. E., & Abualigah, L. (2022). Dwarf mongoose optimization algorithm. *Computer Methods in Applied Mechanics and Engineering*, *391*, 114570. https://doi.org/10.1016/j.cma.2022.114570

Ahmad, B., Jian, W., Ali, Z. A., Tanvir, S., & Ali Khan, M. S. (2019). Hybrid anomaly detection by using clustering for wireless sensor network. *Wireless Personal Communications*, *106*(4), 1841–1853. https://doi.org/10.1007/s11277-018-5721-6

Alghamdi, M. I. (2022). A hybrid model for intrusion detection in IoT applications. *Wireless Communications and Mobile Computing*, *2022*. https://doi.org/10.1155/2022/4553502

Alsaedi, N., Hashim, F., Sali, A., & Rokhani, F. Z. (2017). Detecting Sybil attacks in clustered wireless sensor networks based on energy trust system (ETS). *Computer Communications*, *110*, 75–82. https://doi.org/10.1016/j.comcom.2017.05.006

Anbarasan, M., Muthu, B. A., Sivaparthipan, C. B., Sundarasekar, R., Kadry, S., Krishnamoorthy, S., Samuel, D. J., & Antony Dasel, A. (2020). Detection of flood disaster system based on IoT, Big data and convolutional deep neural network. *Computer Communications*, *150*(November 2019), 150–157. https://doi.org/10.1016/j.comcom.2019.11.022

Bansal, A., & Kaur, S. (2018). *Extreme gradient boosting based tuning for classification in intrusion detection systems. Vol. 905*. Springer Singapore.

Biswas, P., Charitha, R., Gavel, S., & Raghuvanshi, A. S. (2019). Fault detection using hybrid of KF-ELM for wireless sensor networks. *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI*, *2019*(Icoei), 746–750. https://doi.org/10.1109/ICOEI.2019.8862687

Blywis, B. (2009). A real-time and energy-efficient MAC protocol for wireless sensor networks. *International Journal of Ultra Wideband Communications and Systems*, *1*(2), 128–142. https://doi.org/10.1504/IJUWBCS.2009.029002

Cao, B., Li, C., Song, Y., Qin, Y., & Chen, C. (2022). Applied Sciences Network Intrusion Detection Model Based on CNN and GRU.

Cepheli, Ö., Büyükçorak, S., & Karabulut Kurt, G. (2016). Hybrid intrusion detection system for DDoS attacks. *Journal of Electrical and Computer Engineering*, *2016*. https://doi.org/10.1155/2016/1075648

Choi, W., Shah, P., & Das, S. K. (2004). A framework for energy-saving data gathering using Two-phase clustering in wireless sensor networks. *Proceedings of MOBIQUITOUS 2004 - 1st Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*, 203–212. https://doi.org/10.1109/MOBIQ.2004.1331727

Constraints, D. (2016). Source anonymity in WSNs against Global Adversary Utilizing Low Transmission Rates With. https://doi.org/10.3390/s16070957.

Das, S., & Namasudra, S. (2022). A novel hybrid encryption method to secure healthcare data in IoT-enabled healthcare infrastructure. *Computers and Electrical Engineering*, *101*(September 2021), 107991. https://doi.org/10.1016/j.compeleceng.2022.107991

Davahli, A., Shamsi, M., & Abaei, G. (2020). Hybridizing genetic algorithm and grey wolf optimizer to advance an intelligent and lightweight intrusion detection system for IoT wireless networks. *Journal of Ambient Intelligence and Humanized Computing*, *11*(11), 5581–5609. https://doi.org/10.1007/s12652-020-01919-x

Deepa, C., & Latha, B. (2019). HHSRP: A cluster based hybrid hierarchical secure routing protocol for wireless sensor networks. *Cluster Computing*, *22*(s5), 10449–10465. https://doi.org/10.1007/s10586-017-1065-3

Devi, P. P., & Jaison, B. (2020). Protection on wireless sensor network from clone attack using the SDN-enabled hybrid clone node detection mechanisms. *Computer Communications*, *152*(January), 316–322. https://doi.org/10.1016/j.comcom.2020.01.064

Elsaid, S. A., & Albatati, N. S. (2020). An optimized collaborative intrusion detection system for wireless sensor networks. *Soft Computing*, *24*(16), 12553–12567. https://doi.org/10.1007/s00500-020-04695-0

Ezugwu, A. E., Agushaka, J. O., Abualigah, L., Mirjalili, S., & Gandomi, A. H. (2022). *Prairie Dog optimization algorithm. Vol. 34*. Springer.

Farooqi, A. H., Khan, F. A., Wang, J., & Lee, S. (2013). A novel intrusion detection framework for wireless sensor networks. *Personal and Ubiquitous Computing*, *17*(5), 907–919. https://doi.org/10.1007/s00779-012-0529-y

Faysal, J. A., Mostafa, S. T., Tamanna, J. S., Mumenin, K. M., Arifin, M. M., Awal, M. A., Shome, A., & Mostafa, S. S. (2022). XGB-RF: A hybrid machine learning approach for IoT intrusion detection. *Telecom*, *3*(1), 52–69. https://doi.org/10.3390/telecom3010003

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. *Automated Machine Learning: Methods, Systems, Challenges*, 3–33. https://doi.org/10.1007/978-3-030-05318-5_1

Gandhimathi, L., & Murugaboopathi, G. (2020). A novel hybrid intrusion detection using flow-based anomaly detection and cross-layer features in wireless sensor network. *Automatic Control and Computer Sciences*, *54*(1), 62–69. https://doi.org/10.3103/S0146411620010046

Gao, P. (2014). Mechatronics and automatic control systems. *Lecture Notes in Electrical Engineering*, *237*, 1063–1071. https://doi.org/10.1007/978-3-319-01273-5_120

Gebreyesus, G. (2021). Secure intrusion detection system for hierarchically distributed wireless sensor networks. 9–14.

Ghugar, U., Pradhan, J., Bhoi, S. K., & Sahoo, R. R. (2019). LB-IDS: Securing wireless sensor network using protocol layer trust-based intrusion detection system. *Journal of Computer Networks and Communications*, *2019*. https://doi.org/10.1155/2019/2054298

Godala, S., & Vaddella, R. P. V. (2020). A study on intrusion detection system in wireless sensor networks. *International Journal of Communication Networks and Information Security*, *12*(1), 127–141.

Gupta, S. K., Tripathi, M., & Grover, J. (2022). Hybrid optimization and deep learning based intrusion detection system. *Computers and Electrical Engineering*, *100*(May), 107876. https://doi.org/10.1016/j.compeleceng.2022.107876

Han, L., Zhou, M., Jia, W., Dalil, Z., & Xu, X. (2019). Intrusion detection model of wireless sensor networks based on game theory and an autoregressive model. *Information Sciences*, *476*, 491–504. https://doi.org/10.1016/j.ins.2018.06.017

Intelligence, A., Science, S., Media, B., & Nature, S. (2019). A new hybrid approach for intrusion detection using machine learning. *Applied Intelligence*, 2735–2761.

Ismail, A., & Amin, R. (2019). Malicious cluster head detection mechanism in wireless sensor networks. *Wireless Personal Communications*, *108*(4), 2117–2135. https://doi.org/10.1007/s11277-019-06512-w

Jatti, S. A. V., & Kishor Sontif, V. J. K. (2019). Intrusion detection systems. *International Journal of Recent Technology and Engineering*, *8*(2 Special Issue 11), 3976–3983. https://doi.org/10.35940/ijrte.B1540.0982S1119

Jiang, S., Zhao, J., & Xu, X. (2020). SLGBM: An intrusion detection mechanism for wireless sensor networks in smart environments. *IEEE Access*, *8*, 169548–58. https://doi.org/10.1109/ACCESS.2020.3024219

Kanna, P. R., & Santhi, P. (2022). Hybrid intrusion detection using MapReduce based black widow optimized convolutional long short-term memory neural networks. *Expert Systems with Applications*, *194*(October 2021), 116545. https://doi.org/10.1016/j.eswa.2022.116545

Kasongo, S. M. (2021). An advanced intrusion detection system for IIoT based on GA and tree based algorithms. *IEEE Access*, *9*, 113199–113212. https://doi.org/10.1109/ACCESS.2021.3104113

Khilar, R., Mariyappan, K., Christo, M. S., Amutharaj, J., Anitha, T., Rajendran, T., & Batu, A. (2022). Artificial intelligence-based security protocols to resist attacks in Internet of things. *Wireless Communications and Mobile Computing*, *2022*. https://doi.org/10.1155/2022/1440538

Kumar, S. M. (2022). Hybrid optimized deep neural network with enhanced conditional random field based intrusion detection on wireless sensor network. *Neural Processing Letters*, https://doi.org/10.1007/s11063-022-10892-9

Lai, G. H. (2016). Detection of wormhole attacks on IPv6 mobility-based wireless sensor network. *Eurasip Journal on Wireless Communications and Networking*, *2016*(1), https://doi.org/10.1186/s13638-016-0776-0

Li, L., Yu, Y., Bai, S., Cheng, J., & Chen, X. (2018). Towards effective network intrusion detection: A Hybrid Model Integrating Gini Index and GBDT with PSO. 2018.

Liu, W., Cheng, J., Wang, X., Lu, X., & Yin, J. (2022). Hybrid differential privacy based federated learning for Internet of things. *Journal of Systems Architecture*, *124*(July 2021), 102418. https://doi.org/10.1016/j.sysarc.2022.102418

Mahajan, S., Harikrishnan, R., & Kotecha, K. (2022). Prediction of network traffic in wireless Mesh networks using hybrid deep learning model. *IEEE Access*, *10*, 7003–7015. https://doi.org/10.1109/ACCESS.2022.3140646

Mahbooba, B., Timilsina, M., Sahal, R., & Serrano, M. (2021). Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model. *Complexity*, *2021*. https://doi.org/10.1155/2021/6634811

Mangrulkar, R. S., & Negandhi, P. D. (2018). Applications of machine learning in wireless sensor networks. *Soft Computing in Wireless Sensor Networks*, 51–74. https://doi.org/10.1201/9780429438639-3

Meena, G., & Choudhary, R. R. (2017). A review paper on IDS classification using KDD 99 and NSL KDD dataset in WEKA. *2017 International Conference on Computer, Communications and Electronics, COMPTELIX*, *2017*, 553–558. https://doi.org/10.1109/COMPTELIX.2017.8004032

Mohd, N., Singh, A., & Bhadauria, H. S. (2020). A novel SVM based IDS for distributed denial of sleep strike in wireless sensor networks. *Wireless Personal Communications*, *111*(3), 1999–2022. https://doi.org/10.1007/s11277-019-06969-9

Mojtaba, S., Bamakan, H., Wang, H., Yingjie, T., & Shi, Y. (2016). Neurocomputing An effective intrusion detection framework based on MCLP / SVM optimized by time-varying chaos particle swarm optimization. *Neurocomputing*, *199*, 90–102. https://doi.org/10.1016/j.neucom.2016.03.031

Moon, P. S., & Ingole, P. K. (2015). An overview on: Intrusion detection system with secure hybrid mechanism in wireless sensor network. *Conference Proceeding - 2015 International Conference on Advances in Computer Engineering and Applications, ICACEA*, *2015*, 272–277. https://doi.org/10.1109/ICACEA.2015.7164714

Moulad, L., Belhadaoui, H., & Rifi, M. (2019). Implementation of an hierarchical hybrid intrusion detection mechanism in wireless sensor network based on energy management. *Advances in Intelligent Systems and Computing*, *756*, 360–377. https://doi.org/10.1007/978-3-319-91337-7_33

Nancy, P., Muthurajkumar, S., Ganapathy, S., Santhosh Kumar, S. V. N., Selvi, M., & Arputharaj, K. (2020). Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks. *IET Communications*, *14*(5), 888–895. https://doi.org/10.1049/iet-com.2019.0172

Ovelade, O. N., & Ezugwu, A. E. (2021). Ebola optimization search algorithm: A New nature-inspired metaheuristic algorithm for global optimization problems. *International Conference on Electrical, Computer, and Energy Technologies, ICECET*, *2021*(December), 9–10. https://doi.org/10.1109/ICECET52533.2021.9698813

Pajila, P. J. B., Julie, E. G., & Robinson, Y. H. (2021). *FBDR-Fuzzy Based DDoS attack detection and recovery mechanism for wireless sensor networks*. Springer US.

Paliwal, P., & Kumar, D. (2018). *ABC Based Neural Network Approach for Churn Prediction in Telecommunication Sector*. Vol. 84.

Pande, S., Khamparia, A., & Gupta, D. (2021). Feature selection and comparison of classification algorithms for wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing (0123456789)*, https://doi.org/10.1007/s12652-021-03411-6

Panigrahi, R., Borah, S., Pramanik, M., Bhoi, A. K., Barsocchi, P., Nayak, S. R., & Alnumay, W. (2022). Intrusion detection in cyber–physical environment using hybrid naïve Bayes—decision table and multi-objective evolutionary feature selection. *Computer Communications*, *188*(September 2021), 133–144. https://doi.org/10.1016/j.comcom.2022.03.009

Praveen Kumar, D., Amgoth, T., & Annavarapu, C. S. R. (2019). Machine learning algorithms for wireless sensor networks: A survey. *Information Fusion*, *49*(April 2018), 1–25. https://doi.org/10.1016/j.inffus.2018.09.013

Pundir, S., Wazid, M., Singh, D. P., Das, A. K., Rodrigues, J. J. P. C., & Park, Y. (2020). Intrusion detection protocols in wireless sensor networks integrated to Internet of things deployment: Survey and future challenges. *IEEE Access*, *8*, 3343–3363. https://doi.org/10.1109/ACCESS.2019.2962829

Qi, G., Zhou, J., Jia, W., Liu, M., Zhang, S., & Xu, M. (2021). Intrusion detection for network based on elite clone artificial Bee colony and back propagation neural network. *Wireless Communications and Mobile Computing*, *2021*. https://doi.org/10.1155/2021/9956371

Rabbani, M., Wang, Y. L., Khoshkangini, R., Jelodar, H., Zhao, R., & Hu, P. (2020). A hybrid machine learning approach for malicious behaviour detection and recognition in cloud computing. *Journal of Network and Computer Applications*, *151*(May 2019), 102507. https://doi.org/10.1016/j.jnca.2019.102507

Regan, R., & Leo Manickam, J. M. (2019). An optimized energy saving model for hybrid security protocol in WMN. *National Academy Science Letters*, *42*(6), 489–501. https://doi.org/10.1007/s40009-019-0789-4

Ren, J., Guo, J., Qian, W., Yuan, H., Hao, X., & Jingjing, H. (2019). Building an effective intrusion detection system by using hybrid data optimization based on machine learning algorithms. *Security and Communication Networks*, *2019*. https://doi.org/10.1155/2019/7130868

Reshma, V. K., Khan, I. R., Niranjanamurthy, M., Aggarwal, P. K., Hemalatha, S., Almuzaini, K. K., & Amoatey, E. T. (2022). Hybrid block-based lightweight machine learning-based predictive models for quality preserving in the Internet of Things- (IoT-) based medical images with diagnostic applications. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/8173372

Rose, T., Kifayat, K., Abbas, S., & Asim, M. (2020). A hybrid anomaly-based intrusion detection system to improve time complexity in the Internet of energy environment. *Journal of Parallel and Distributed Computing*, *145*, 124–139. https://doi.org/10.1016/j.jpdc.2020.06.012

Rouissi, N., Gharsellaoui, H., & Bouamama, S. (2019). Improvement of watermarking-LEACH algorithm based on trust for wireless sensor networks. *Procedia Computer Science*, *159*, 803–813. https://doi.org/10.1016/j.procs.2019.09.239

Roy, P., & Chowdhury, C. (2021). A survey of machine learning techniques for indoor localization and navigation systems. *Journal of Intelligent and Robotic Systems: Theory and Applications*, *101*(3), https://doi.org/10.1007/s10846-021-01327-z

Sadikin, F., van Deursen, T., & Kumar, S. (2020). A ZigBee intrusion detection system for IoT using secure and efficient data collection. *Internet of Things*, *12*, 100306. https://doi.org/10.1016/j.iot.2020.100306

Saheed, Y. K., Abdulganiyu, O. H., Tchakoucht, T. A., & Rakshit, S. (2022). A Novel Wrapper and Filter-Based Feature Dimensionality Reduction Methods for Anomaly Intrusion Detection in Wireless Sensor Networks. 1–22.

Saidi, A., Benahmed, K., & Seddiki, N. (2020). Secure cluster head election algorithm and misbehavior detection approach based on trust management technique for clustered wireless sensor networks. *Ad Hoc Networks*, *106*. https://doi.org/10.1016/j.adhoc.2020.102215

Saif, S., Das, P., Biswas, S., Khari, M., & Shanmuganathan, V. (2022). HIIDS: Hybrid intelligent intrusion detection system empowered with machine learning and metaheuristic algorithms for application in IoT based healthcare. *Microprocessors and Microsystems*, *104622*. https://doi.org/10.1016/j.micpro.2022.104622

Sakthivel, T., & Chandrasekaran, R. M. (2018). A dummy packet-based hybrid security framework for mitigating routing misbehavior in multi-Hop wireless networks. *Wireless Personal Communications*, *101*(3), 1581–1618. https://doi.org/10.1007/s11277-018-5778-2

Saravana Kumar, N. M., Suryaprabha, E., & Hariprasath, K. (2021). Machine learning based hybrid model for energy efficient secured transmission in wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing*, https://doi.org/10.1007/s12652-021-02946-y

Shi, L., & Li, K. (2022). Privacy protection and intrusion detection system of wireless sensor network based on artificial neural network. *Computational Intelligence and Neuroscience*, *2022*, 1795454. https://doi.org/10.1155/2022/1795454

Singh, A., Amutha, J., Nagar, J., & Sharma, S. (2023). A deep learning approach to predict the number of K-barriers for intrusion detection over a circular region using wireless sensor networks. *Expert Systems with Applications*, *211*(July 2021), 118588. https://doi.org/10.1016/j.eswa.2022.118588

Singh, A., Amutha, J., Nagar, J., Sharma, S., & Lee, C. C. (2022a). AutoML-ID: Automated machine learning model for intrusion detection using wireless sensor network. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-021-99269-x

Singh, A., Amutha, J., Nagar, J., Sharma, S., & Lee, C. C. (2022b). LT-FS-ID: Log-transformed feature learning and feature-scaling-based machine learning algorithms to predict the k-barriers for intrusion detection using wireless sensor network. *Sensors*, *22*(24), 3. https://doi.org/10.1109/JSEN.2022.3226962

Singh, A., Nagar, J., Sharma, S., & Kotiyal, V. (2021). A Gaussian process regression approach to predict the K-barrier coverage probability for intrusion detection in wireless sensor networks. *Expert Systems with Applications*, *172*(October 2020), 114603. https://doi.org/10.1016/j.eswa.2021.114603

Singh, R., Singh, J., & Singh, R. (2017). Fuzzy based advanced hybrid intrusion detection system to detect malicious nodes in wireless sensor networks. *Wireless Communications and Mobile Computing*, *2017*. https://doi.org/10.1155/2017/3548607

Suleiman, M. F., & Issac, B. (2018). Performance comparison of intrusion detection machine learning classifiers on benchmark and new datasets. *28th International Conference on Computer Theory and Applications, ICCTA 2018 - Proceedings*, 19–23. https://doi.org/10.1109/ICCTA45985.2018.9499140

Sun, P., Liu, P., Li, Q., Liu, C., Lu, X., Hao, R., & Chen, J. (2020). DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system. *Security and Communication Networks*, *2020*. https://doi.org/10.1155/2020/8890306

Swarna Priya, R. M., Maddikunta, P. K. R., Parimala, M., Koppu, S., Gadekallu, T. R., Chowdhary, C. L., & Alazab, M. (2020). An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture. *Computer Communications*, *160*(May), 139–149. https://doi.org/10.1016/j.comcom.2020.05.048

Ullah, S., Khan, M. A., Ahmad, J., Jamal, S. S., Huma, Z. E., Hassan, M. T., Nikolaos Pitropakis, A., & Buchanan, W. J. (2022). HDL-IDS: A hybrid deep learning architecture for intrusion detection in the Internet of vehicles. *Sensors*, *22*(4), 1–20. https://doi.org/10.3390/s22041340

Umarani, C., & Kannan, S. (2020). Intrusion detection system using hybrid tissue growing algorithm for wireless sensor network. *Peer-to-Peer Networking and Applications*, *13*(3), 752–761. https://doi.org/10.1007/s12083-019-00781-9

Upadhyay, D., Manero, J., Zaman, M., & Sampalli, S. (2021). Learning classifiers for intrusion detection on power grids. *Ieee Transactions on Network and Service Management*, *18*(1), 1104–1116. https://doi.org/10.1109/TNSM.2020.3032618

VenkataRao, S., & Ananth, V. (2021). A hybrid optimization algorithm and shamir secret sharing based secure data transmission for IoT based WSN. *International Journal of Intelligent Engineering and Systems*, *14*(6), 498–506. https://doi.org/10.22266/ijies2021.1231.44

Vinitha, A., Rukmini, M. S. S., & Dhirajsunehra. (2019). Secure and energy aware multi-Hop routing protocol in WSN using Taylor-based hybrid optimization algorithm. *Journal of King Saud University - Computer and Information Sciences (xxxx)*. https://doi.org/10.1016/j.jksuci.2019.11.009.

Wang, Y., Ma, J., Sharma, A., Singh, P. K., Gaba, G. S., Masud, M., & Baz, M. (2021). An exhaustive research on the application of intrusion detection technology in computer network security in sensor networks. *Journal of Sensors*, *2021*. https://doi.org/10.1155/2021/5558860

Wu, Y., Wei, D., & Feng, J. (2020). Network attacks detection methods based on deep learning techniques: A survey. *Security and Communication Networks*, *2020*. https://doi.org/10.1155/2020/8872923

Yang, L., Moubayed, A., & Shami, A. (2022). MTH-IDS: A multitiered hybrid intrusion detection system for Internet of vehicles. *IEEE Internet of Things Journal*, *9*(1), 616–632. https://doi.org/10.1109/JIOT.2021.3084796

Yin, C., Zhang, S., Yin, Z., & Wang, J. (2019). Anomaly detection model based on data stream clustering. *Cluster Computing*, *22*(s1), 1729–1738. https://doi.org/10.1007/s10586-017-1066-2

Yuan, Y., Huo, L., Wang, Z., & Hogrefe, D. (2018). Secure APIT localization scheme against Sybil attacks in distributed wireless sensor networks. *IEEE Access*, *6*, 27629–27636. https://doi.org/10.1109/ACCESS.2018.2836898

Zhang, W., Han, D., Li, K. C., & Massetto, F. I. (2020). Wireless sensor network intrusion detection system based on MK-ELM. *Soft Computing*, *24*(16), 12361–12374. https://doi.org/10.1007/s00500-020-04678-1

Zhao, G., Wang, Y., & Wang, J. (2023). Lightweight intrusion detection model of the Internet of Things with Hybrid Cloud-Fog Computing. 2023.

Zou, Y., Zhu, J., Wang, X., & Hanzo, L. (2016). A survey on wireless security: Technical challenges, recent advances, and future trends. *Proceedings of the IEEE*, *104*(9), 1727–1765. https://doi.org/10.1109/JPROC.2016.2558521

Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023)
IEEE Xplore Part Number: CFP23AWO-ART; ISBN: 979-8-3503-9663-8

Design and Analysis of 4*4-Bit Vedic Multiplier with 32nm CNFET Technology

Ritik Gupta[1], Rohit Ghughtyal[1], Sai Prateek Mahapatra[1], Sonal Singh[1*]
[1*]Department of Electronics and Communication Engineering,
Delhi Technological University,
New Delhi, India
*Corresponding author(s). E-mail(s): sonal.singh0811@gmail.com;
Contributing authors: ritik2001gupta@gmail.com, rohitghughtyal96@gmail.com, saiprateekmahapatra@gmail.com;

2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192881

***Abstract*— Carbon Nanotubes Field Effect Transistors (CNFETs) have shown numerous advantages over conventional Metal Oxide Semiconductor FETs (MOSFETs) and CMOS in terms of decreased short channel effect, lower power consumption and high performance. As the transistor channel length shortens to below 10 nm in length, short channel effects, such as drain-induced barrier lowering and velocity saturation, becomes prominent, which negatively impact transistor performance and reliability. As CNFETs do not show scaling issues, CNFETs based digital circuit designs, such as adder and multiplier, can possibly show better results in terms of power consumption and propagation delay. Multiplication operation is a crucial operation whose application ranges from computing convolution to fast fourier transform. Hence, it is important to have an efficient architecture and high-performance multiplier. Array multipliers, Wallace and Booth's multipliers are some of the existing multipliers. One approach to perform high-performance multiplication is through the Vedic Multiplier, whose algorithm reduces delay and power consumption. Many of the pre-existing Vedic multipliers implementations techniques such as MUX based adder, Carry select adder, Hybrid 13T Full Adder rely on CMOS technology for their implementation which brings the challenges like short channel effect while scaling of CMOS transistors. Because CNFETs do not show scaling issues the performance of Vedic multiplier can be improved using CNFETs. This paper proposes a 4*4-bit Vedic Multiplier using 32nm CNFET Technology. The proposed implementation is simulated using HSPICE simulation tool, CosmosScope (cscope) software and Stanford CNFET Model file. The delay and power consumption parameters are measured for various supply voltages. Moreover, the comparative analysis of the aforementioned parameters has also been done by changing the chirality of the transistors.**

Keywords*— *Carbon Nanotubes Field Effect Transistors, Vedic multiplication, Urdhva Tiryagbhyam, Chiral Vector

I. INTRODUCTION

Carbon Nanotubes Field Effect Transistors (CNFETs) have developed as potential replacement to the conventional Metal Oxide Semiconductor FETs because of scaling down related issues which are seen in MOSFET. Some of the issues that are confronted with MOSFETs during scaling down are short channel effects, hot carrier generation, drain-induced barrier lowering, high leakage current and velocity saturation [1]. In the area of having decreased short channel effect, lesser area, lower power consumption and higher performance, CNFETs show better performance than CMOS [2]. Along with CNFETs, FinFET, single-electron transistor (SET), and tunnel FET have been made to solve the issues faced with conventional MOSFET and CMOS. MOSFETs depend on the phenomenon of diffusion and drift whereas CNFETs, which have high carrier velocity, are based on the ballistic transport conduction mechanism [1]. While, in MOSFET, the channel material is Si-based, CNFET have their channel material constructed using carbon. It can be observed that thinner structures of CNFETs support better gate control in comparison with conventional MOSFET.

Multiplication is one of the most fundamental operations in any mathematical operation, complex or simple. It is widely used in Arithmetic Logic Units, and in Embedded Systems, like in Digital Signal Processors. Operation like Multiplication is one such operation which can decide the overall execution time for any processing unit. Hence, building an efficient multiplier can have important applications in computation. Operations involving convolution and fast fourier transform, which have their own importance, are primarily based on multiplication. There exist many architectures which have been developed to increase the efficiency of multiplication, for instance, array multipliers, Wallace Multipliers, Booth's multipliers, etc.

Another approach to perform multiplication is with the help of Vedic Multipliers. A Vedic Multiplier has better performance than conventional multiplier [3]. The term 'Vedic' basically comes out from the root word "Veda", signifying storage of data and information. Vedic mathematics is an interesting as well as important field of study. It is useful in reducing the steps for obtaining the summation or multiplication results as it offers the easiest steps in calculations. Over the years, there have been several Vedic multipliers which have reduced the input to output delay and enhanced area and power efficiency, by improving the circuit as well as by using different Transistor technologies. Some of the previous Vedic multiplication implementation include using MUX based adder [3], carry select adder [4], GDI logic [5], Manchester carry chain (MCC) adder [6], hybrid 13T full adder [7], Brent Kung Adders [8], etc. [3] uses full adder based on MUX which improves speed of Vedic multiplier, [5] uses a reduced transistor approach to implement Vedic multiplier, [6] shows better power-delay product (PDP) than existing designs with minimum area and [7] reduces power consumption and delay using 90nm CMOS technology. The demerits of existing Vedic multipliers is that it is based on CMOS technology which raises challenges like short channel effect while scaling of CMOS transistors. As the various advantages of CNFET have been previously discussed, through this paper an approach has been presented to implement Vedic multipliers using CNFETs. As CNFETs have significant merits over other transistors such as lesser short channel effect, lower power consumption, the proposed approach to implement CNFET based Vedic multiplier is expected to show a more preferable result. As per our literature survey and to the best of our knowledge, previous implementation of CNFET based Vedic multiplier is unexplored. As the novelty, the paper proposes architecture of CNFET based Vedic multiplier.

In this paper, we have implemented a 32 nm CNFET based 4*4-bit Vedic multiplier and various analysis such as delay and power consumption have been found via utilizing the

Authorized licensed use limited to: DELHI TECHNICAL UNIVERSITY. Downloaded on August 10,2023 at 06:40:09 UTC from IEEE Xplore. Restrictions apply.

HSPICE simulation tool, CosmosScope (cscope) and CNFET model developed at Stanford University [9]. Section II describes about CNFETs and Section III presents algorithm for Vedic multiplication. Section IV shows proposed methodology and includes measured results like power consumption and delay for implemented multiplier circuit.

II. Carbon Nanotubes Field Effect Transistors (CNFETs)

CNFETs have been developed as an improvement to conventional transistors [10-11]. The fabrication of CNFETs is done using multiple Carbon Nanotubes (CNTs), which are used as channel in the transistor. Techniques like Chemical Vapor Deposition (CVD), LASER ablation and Arc Discharge are used for developing the CNTs. Various fabrication processes like doping, masking, etching and lithography help in CNFETs' development [1]. Metal deposition is then carried out to connect the CNTs with the gate, source and drain regions of the CNFET.

In CNFETs, which are primarily based on the phenomenon of ballistic transportation, the channel length is a factor responsible for such transportation because ballistic transportation is possible when the channel length of CNFET is lesser than mean free path of charge carrier [12]. Carbon nanotubes or CNTs, which form the channel material of CNFETs, are responsible for favorable features possessed by CNTFETs such as high mechanical strength, lower power consumption, high electrical conductivity and reduced size which is suitable for nanoelectronics. These properties of CNTs improves overall performance of any digital circuit made using CNFETs. CNTs can be categorized on the basis of two parameters, namely chirality and number of walls.

On the basis of the walls, CNTs are of two kinds : single walled (SWCNT) and multiwalled (MWCNT). For the manufacturing of CNFETs, single-walled are preferred over multi-walled because it exhibits better semiconductor properties vis–a-vis multiwalled, and also has the capability of withstanding the large flow of current.

Chirality is a unique property in CNTs which defines its electrical, optical and mechanical characteristics [13]. It is explained as the orientation in which graphene, which has a hexagonal lattice of carbon atoms, is arranged to form the wall of the carbon nanotube. For SWCNTs, this orientation is described by a set of pair of integers (n,m). The coordinates of the vector gives us the direction of the graphene lattice that is rolled up to form the wall of the carbon nanotube [14]. Moreover, the chirality of the nanotube also determines its bandgap of electrons, thus, making it an important factor to be considered when considering the design and engineering efficiency of carbon nanotube-based materials. On the basis of chirality or chiral vector(n,m), CNTs can have three varieties : zig-zag, helical and armchair [1]. Armchair CNTs have chiral vectors that are of the form (n,n), while zigzag CNTs have chiral vectors of the form (n,0) or (0,n), where n is an integer. Helical chirality in CNTs refers to the twisting of the graphene lattice as it is rolled up into a tube along the length of the tube. Usually, zig-zag type CNTs have a larger bandgap as compared to armchair type, due to chiral symmetry. Any CNT with indices (n, m) is metallic on the following condition: n=m or n-m = 3i (where, i is an integer). Any CNT with indices (n, m) is semiconducting on the following condition: n-m ≠3i (where, i is an integer). CNFETs are developed using semiconducting CNTs. Using the SWCNT chiral vector (n,m) can help in determining its diameter as follows [15-16]:

$$d_{CNT} = \frac{\sqrt{3}a_{c-c}}{\pi}\sqrt{m^2 + n^2 + mn} \quad \text{... (i)}$$

where a_{c-c} is the carbon-carbon bond length, whose value is about 1.42 Å.

Chirality plays a significant role in creating CNFETs and can impact their performance. A CNFET is a type of transistor that uses a carbon nanotube as the conducting channel, with a gate electrode controlling the flow of current through the channel. CNTs with different chiral vectors have different bandgap energies, which can affect the threshold voltage of a CNFET [17]. Chirality also affects the conductivity of CNFETs [18]. Since semiconducting CNTs have moderate bandgaps, they are ideal for their use as transistors. Whereas, metallic CNTs have lower resistance, which can be used as conducting channels in the same CNFET. CNFETs can have either a p-channel (PCNFET) or a n-channel (NCNFET). Keeping the Chirality at (19,0), Drain Current-Drain to Source Voltage (I_D-V_{DS}) Characteristics is plotted for an NCNFET, for different gate-source voltages (V_{GS}), shown in Fig. 1.

As described above, CNFET is a field Effect Transistor built out of Carbon Nanotube. CNFETs have several advantages over traditional silicon-based transistors, including low power consumption and improved thermal stability [17]. Its high electron mobility, or the velocity of electrons through the channel, compared to traditional silicon-based transistors, is higher, which allows for faster switching times and lower power consumption [19]. They can also withstand higher temperatures without thermal degradation [17]. Moreover, as the name 'carbon nanotube' suggests, they are quite small in diameter and thus can be used extensively in areas which have severe constraints on the size of appliance, such as nanoelectronics.

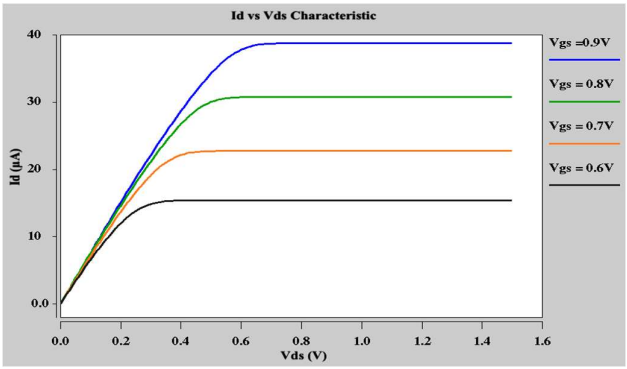


Fig. 1. I_D-V_{DS} characteristics of an NCNFET

Chirality also affects the optical properties of CNTs, like its absorption and emission spectra. Thus, controlling these properties allows researchers to develop newer technologies in optoelectronics applications [20], healthcare applications [21-23] , in photovoltaic devices [24]. Of late, researchers have also developed various methods for synthesizing CNTs of specific chirality to suit the application-specific demand.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:40:09 UTC from IEEE Xplore. Restrictions apply.

III. Algorithm For Vedic Multiplication

The initial references to the Vedic Multiplication techniques were documented in [25] with other 16 principles and concepts, of which the Urdhva Tiryagbhyam (UT) Sutra is one of the important principles concerned with multiplication. The UT Sutra is one of the oldest in Vedic mathematics history, applicable to various techniques and applications in math, as it holds in all situations of multiplications. The algorithm depends on the idea, "vertical and crosswise" multiplication. It is recursive for complex or high no. of bits multiplications, with base condition at 2*2 bit. For base condition, the LSB bits of the two numbers are multiplied vertically, then the remaining corresponding non-MSB bits (2 bits) are multiplied crosswise and added. Then, the MSB bits are vertically multiplied. The result at the end of each step will be appended to give the result.

For even *n*, consider two *n* - bit numbers, P and Q, which can be divided into *n*/2 bits, say P_H, P_L, Q_H and Q_L respectively. P_H and P_L represent most significant and least significant *n*/2 bits of P. Similarly, Q_H and Q_L represent most significant and least significant *n*/2 bits of Q. Using this we can obtain the following partial product:

$$P_0 = P_L * Q_L \qquad P_1 = P_L * Q_H$$
$$P_2 = P_H * Q_L \qquad P_3 = P_H * Q_H$$

where, P_0 , P_1 , P_2 and P_3 partial products can be added as following to get required result:

$$P * Q = P_3 * 2^n + (P_2 + P_1) * 2^{n/2} + P_0$$

IV. Proposed Architecture

A. Full Adder

A Full adder is a quintessential component of the Vedic Multiplier. From the top-level design, it involves taking 3 sets of 1-bit inputs, while giving out the Sum and Carry generated by adding the 3 data inputs. Consider the inputs as A, B, C_{in} and the results as CARRY and SUM. Then, the Boolean function of the SUM and CARRY as a function of the inputs is given as:

$$\text{SUM} = A \oplus B \oplus C_{in} \quad \dots (1)$$

$$\text{CARRY} = AB + (A \oplus B) C_{in} \quad \dots (2)$$

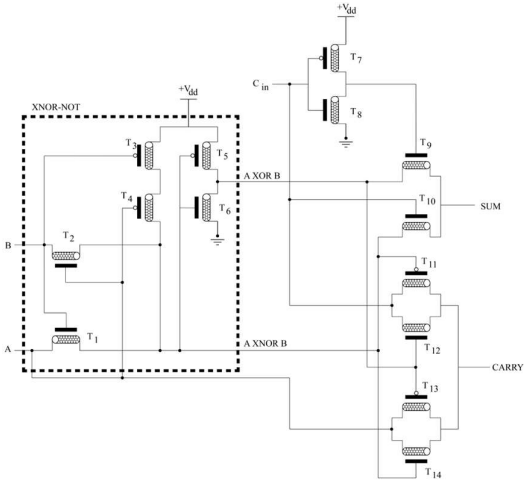


Fig. 2. 14T CNFET 1-bit Full Adder (adapted from [26])

The 1-Bit Full adder used in our architecture, shown in Fig. 2. , works by realizing the design for the boolean expressions in equation (1) & (2). It consists majorly of the XNOR-NOT module followed by the SUM and CARRY operations. The XNOR-NOT module realizes both XOR as well as XNOR expressions. This Module is created using 6 Transistors. Transistors T_1, T_2 (NCNFET)and T_3, T_4 (PCNFET) create the XNOR gate, the output of which is then passed through a CNFET inverter to generate XOR output. The third input (C_{in}) has also been passed through an inverter for the next step. Both XNOR and XOR are used subsequently in the SUM and CARRY Module. Subsequently, the SUM operation can be performed by XORing the 3 data inputs and knowing that the first two data inputs have already been XORed previously, we can use a Pass-Transistor Logic (PTL) to generate the result. The logic behind PTL, besides reducing transistor count, is also to reduce the delay of the SUM module. The total transistors of 2 NCNFET transistors (T_9 & T_{10}) are used in this module.

To generate CARRY, the use of the Transmission Gates Logic (TG) has been taken. As Eqn. (2) denotes, there are two minterms and consequently, there will be two TGs used in the Carry Module – one each to generate the respective minterm, i.e., AB & (A⊕B) C_{in}. Each submodule uses a PCNFET (T_{11} & T_{13}) and NCNFET (T_{12} & T_{14}) transistors respectively, thus making the total of 4 Transistors (T_{11}-T_{14}) in the Carry Module. Hence, the entire 1-Bit CNFET Full Adder is made using 14 Transistors. The full adder's use has been extended to create the other components, thus realizing the Vedic Multiplier's architecture.

B. Proposed 4*4-Bit Vedic Multiplier

The presented architecture is the 4*4-bit Vedic multiplier based on 32nm CNFET Technology. Fig. 3. shows CNFET based 2*2-bit Vedic multiplier which computes the binary multiplication of a pair of 2-bit numbers using Urdhva Tiryagbhyam Sutra. The proposed 4*4-bit Vedic multiplier, as depicted in Fig. 4. is considered as a black box which takes in a pair of 4-bit data as inputs and gives out an 8-bit result. It constitutes CNFET based 2*2 - bit Vedic multiplier and CNFET based Ripple Carry Adders. The CNFET based Ripple Carry adder functions as a serial full adder, which generates sequentially the output bits viz. sum and carry bits. Then, 4-bit Ripple Carry Adder (i) is used to add the partially generated array of SUM bits generated by the 2-bit Vedic Multipliers (ii) and (iii). The result obtained is used to add the result with the result of 2-bit Vedic Multipliers (i) and (iv). The 2-bit Ripple Carry Adder is used to add M3[3:2] with the result of 1-bit Full Adder, i.e., sum of the two carries generated by the two 4-bit Ripple Carry adders. The results are the two most significant bits of the final result. The carry generated in the final step is discarded.

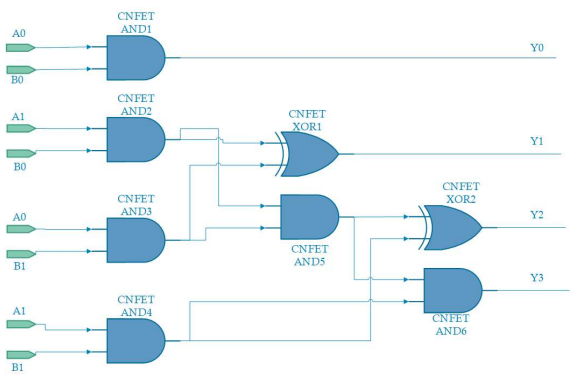


Fig. 3. CNFET based 2*2-bit Vedic multiplier

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:40:09 UTC from IEEE Xplore. Restrictions apply.

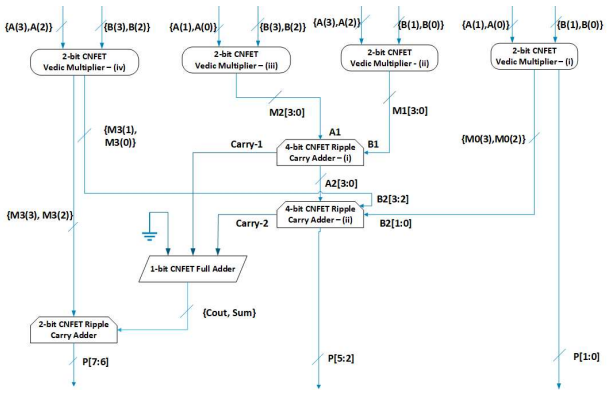


Fig. 4. Proposed 4*4 - bit Vedic Multiplier

The Table 1 elucidates the step-by-step working of the Vedic Multiplier for better understanding of the working and mathematical operations performed to obtain the final result.

TABLE I. COMPONENT-LEVEL OUTPUTS TO EXPLAIN THE FINAL MULTIPLICATION RESULT OF PROPOSED 4*4-BIT CNFET VEDIC MULTIPLIER

Input / Net Name	Binary	Decimal	Remarks
A	1111	15	1st input
B	1111	15	2nd input
M0	1001	9	2-bit CNFET Vedic Multiplier (i) - Output
M1	1001	9	2-bit CNFET Vedic Multiplier (ii) - Output
M2	1001	9	2-bit CNFET Vedic Multiplier (iii) - Output
M3	1001	9	2-bit CNFET Vedic Multiplier (iv) - Output
A2	0010	2	4-bit CNFET Ripple Carry Adder (i)
Carry-1	1	1	Carry of 4-bit CNFET Ripple Carry Adder (i)
A2[3:0] + {M3[1:0],M0[3:2] }	1000	8	4-bit CNFET Ripple Carry Adder (ii)
Carry-2	0	0	Carry of 4-bit CNFET Ripple Carry Adder (ii)
(Cout,S)	(0,1)	1	1-bit CNFET Full Adder (carry 1 + carry 2)
P[7:6]	11	3	2-bit CNFET Ripple Carry Adder Output
P[5:2]	1000	8	4-bit CNFET Ripple Carry Adder (ii)
P[1:0]	01	1	2-bit CNFET Vedic Multiplier (i) - Output

C. Implementation And Result

The proposed 32nm CNFET based Vedic multiplier (Fig. 4) has been implemented using HSPICE and simulated using CosmosScope. The CNFET model file was developed by and resourced from the Nanoelectronics Lab, Stanford University. For the purpose of carrying out a simulation for the 4*4-bit Vedic Multiplier, single walled CNFETs were used and the dimensions for a chiral vector were set to be (19,0).

The reason for keeping the chirality at (19,0) is to have a lower threshold voltage of CNFET, which facilitates in having an increased range of operation. The single-walled CNFETs are used to have a more uniform control over the properties of the transistor, which makes its study with variation with chirality easier. The simulation result of the proposed Vedic Multiplier at a supply voltage of 1V and chiral vector (19,0) is shown in Fig. 5.

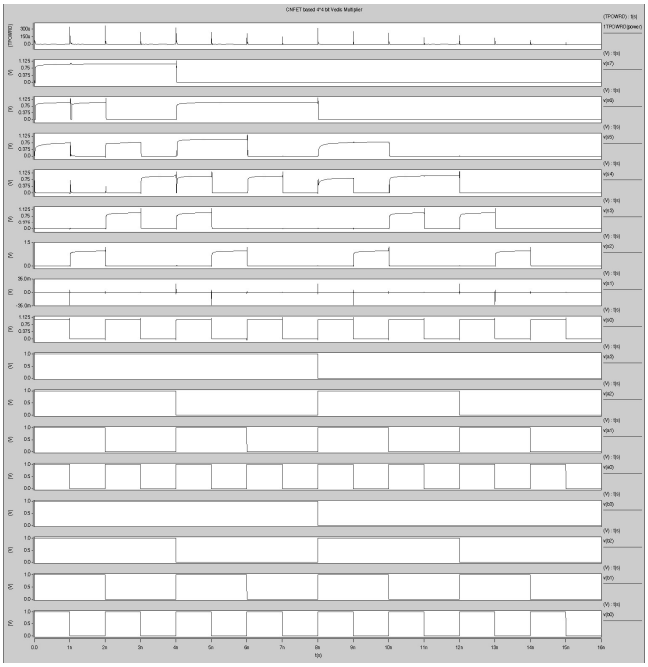


Fig. 5. HSPICE Simulation at supply voltage of 1V

Furthermore, to facilitate a deeper understanding of the proposed Vedic Multiplier, a comparative analysis of the maximum input-to-output delay and power consumption is performed. This is done by varying the supply voltages from 0.6V to 1.0V, by giving a step size of 0.1V. However, the chiral vector is the same as used above, i.e., at (19,0), to ensure that transistor characteristics are kept constant during the simulation. The results obtained are consolidated in Table 2. The trends of the variation in measured values for delay and power consumption are also plotted in Fig. 6. and Fig. 7. respectively. Further, to observe the effect of chirality on overall performance of the proposed circuit, analysis of propagation delay and power consumption is performed by varying chiral vector. Thus, as chirality is varied, the supply voltage(V_{dd}) is kept constant. Chiral vector is varied as, (16,0), (17,0), (19,0), (20,0) and then (22,0). While the chiral vector is varied, the supply voltage (V_{dd}) is first kept constant at 0.7V and then at 0.9V. Table 3 provides the obtained result for supply voltages 0.7V and 0.9V. The trends of the variation in measured values for delay and power consumption in Table 3 are also plotted in Fig. 8. and Fig. 9. respectively.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:40:09 UTC from IEEE Xplore. Restrictions apply.

TABLE II. DELAY AND POWER CONSUMPTION ON SUPPLY VOLTAGE VARIATION

Parameters	Supply Voltage (V)	Result
Delay	0.6	144.265 ps
	0.7	90.7855 ps
	0.8	65.0155 ps
	0.9	52.411 ps
	1.0	44.09 ps
Power Consumption	0.6	0.35093 µW
	0.7	0.52167 µW
	0.8	0.71812 µW
	0.9	0.97419 µW
	1.0	1.3983 µW

TABLE III. DELAY AND POWER CONSUMPTION FOR SUPPLY VOLTAGE 0.7V AND 0.9V

Parameters	Chirality	Result	
		V_{dd} = 0.7V	V_{dd} = 0.9V
Delay	(16,0)	209.2745 ps	90.075 ps
	(17,0)	149.568 ps	73.0515 ps
	(19,0)	90.7855 ps	52.411 ps
	(20,0)	74.0415 ps	45.754 ps
	(22,0)	57.127 ps	38.8395 ps
Power Consumption	(16,0)	0.43926 µW	0.86307 µW
	(17,0)	0.47157 µW	0.92072 µW
	(19,0)	0.52167 µW	0.97419 µW
	(20,0)	0.54522 µW	1.0119 µW
	(22,0)	0.60062 µW	1.1041 µW

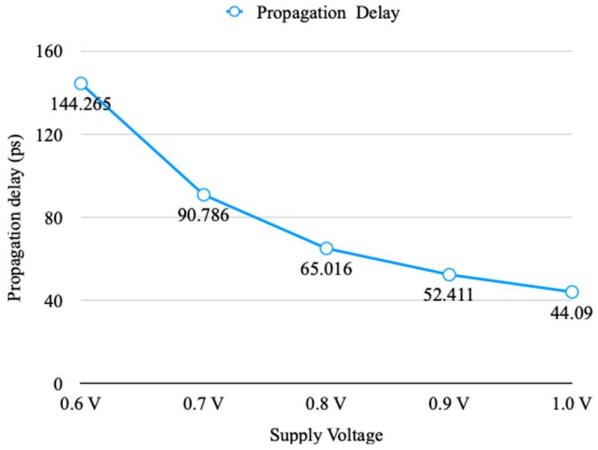


Fig. 6. Propagation Delay vs Supply Voltage

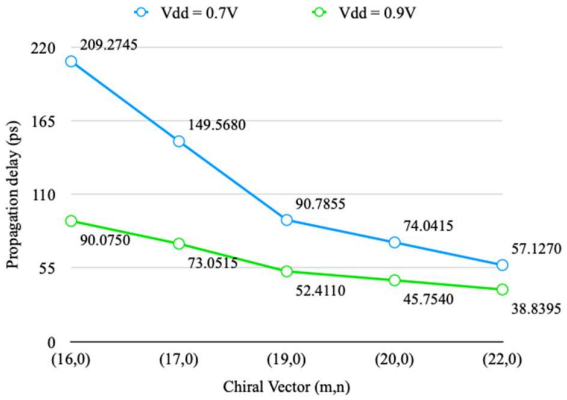


Fig. 8. Propagation delay on varying chiral vector at supply voltages 0.7V and 0.9V

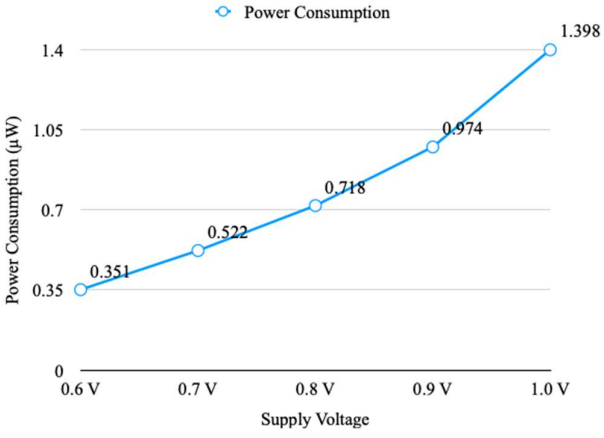


Fig. 7. Power Consumption vs Supply voltage

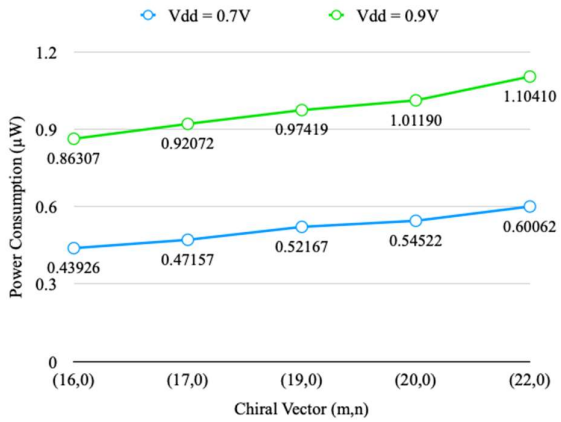


Fig. 9. Power Consumption on varying chiral vector at supply voltages 0.7V and 0.9V

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:40:09 UTC from IEEE Xplore. Restrictions apply.

For the proposed architecture, it is observed that as supply voltage is varied from 0.6V to 1.0V (for chiral vector (19,0)), propagation delay decreases and power consumption increases. As supply voltage increases, faster switching occurs in transistors present in the circuit which results in reduced propagation delay. Also, as chiral vector is varied as (16,0) , (17,0), (19,0), (20,0) and (22,0) and supply voltage is kept at 0.7V and 0.9V respectively, it is observed that propagation delay decreases and power consumption increases.

The proposed model, when simulated at supply voltage of 0.9V and chiral vector (19,0) results in a propagation delay of 52.411ps and power consumption of 0.974µW. Even though based on different technologies, the proposed results can be analyzed by taking into account already existing methods, as presented in [4] using CMOS-based Carry Save Adder wherein the propagation delay is 376ps and power consumption is 4.11µW. Also considering the results obtained in [6] using CMOS-based MCC Adders, the propagation delay is 698ps and power consumption is 39.81µW.

V. Conclusion

In the proposed method, a novel Carbon Nanotube FET (CNFET) based 4*4-bit Vedic Multiplier is implemented using the 32nm Stanford University CNFET Model. The performance of the Vedic Multiplier is tested by simulating it at different supply voltages from 0.6V to 1V (keeping the chirality at (19,0)) and measuring the maximum delay and the power consumed by the circuit. It has been noted that for a supply voltage of 0.6V, the circuit runs with a highest delay since being closer to threshold voltage, it has slow switching speed. Moreover when supply voltage is increased, it is observed that the delay progressively reduces with a depreciating rate of decrease. Between 0.6V and 0.7V, the rate of decrease is ~37%, while the same for 0.7V to 0.8V is 28.4%, 0.8V to 0.9V is 19.4%, and 0.9V to 1V is 15.9% . Similarly, it is seen that with increase in supply voltage, the power consumption also gradually increases, which conforms with the view that increasing power supply leads to power consumed by the circuit. The maximum delay and power consumption by this Vedic Multiplier is also compared on the basis of chirality of CNFET. The chirality (n,m) is varied as (16,0) , (17,0), (19,0), (20,0) and (22,0) for two supply voltages, 0.7V and 0.9V. The trend observed is that while keeping the voltage constant, the delay progressively decreases and power consumed increases with increasing chirality (i.e. for n=16, 17, 19, 20 and 22).

In future, the proposed work can be extended to multiply two numbers of higher bit-width. Moreover, the scope of the proposed architecture can be explored further by analyzing the effect of channel length of CNFET on the performance of Vedic Multiplier.

References

[1] Gupta, N., Dixit, A. (2021). Carbon Nanotube Field-Effect Transistors (CNFETs): Structure, Fabrication, Modeling, and Performance. In: Hazra, A., Goswami, R. (eds) Carbon Nanomaterial Electronics: Devices and Applications. Advances in Sustainability Science and Technology. Springer, Singapore. https://doi.org/10.1007/978-981-16-1052-3_9

[2] A. Keshavarzi, A. Raychowdhury, J. Kurtin, K. Roy and V. De, "Carbon Nanotube Field-Effect Transistors for High-Performance Digital Circuits—Transient Analysis, Parasitics, and Scalability," in IEEE Transactions on Electron Devices, vol. 53, no. 11, pp. 2718-2726, Nov. 2006, doi: 10.1109/TED.2006.883813.

[3] S. M. Antony, S. S. R. Prasanthi, S. Indu and R. Pandey, "Design of high speed Vedic multiplier using multiplexer based adder," 2015 International Conference on Control Communication & Computing India (ICCC), Trivandrum, India, 2015, pp. 448-453, doi: 10.1109/ICCC.2015.7432938.

[4] S. T. Mestry, S. V. Sankpal and D. N. Golatkar, "Low Power High Performance 4 bit Vedic-multiplier in 32nm," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-5, doi: 10.1109/I2CT51068.2021.9417974.

[5] C. Selvakumari, M. Jeyaprakash and A. Kavitha, "Transistor level implementation of a 8 bit multiplier using Vedic mathematics in 180nm technology," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2016, pp. 1514-1520.

[6] R. Katreepalli and T. Haniotakis, "Power-delay-area efficient design of Vedic multiplier using adaptable manchester carry chain adder," 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2017, pp. 1418-1422, doi: 10.1109/ICCSP.2017.8286618.

[7] Lee, S. J., & Ruslan, S. H. (2018). A 2x2 Bit Multiplier Using Hybrid 13T Full Adder with Vedic Mathematics Method. International Journal of Integrated Engineering, 10(3). Retrieved from https://penerbit.uthm.edu.my/ojs/index.php/ijie/article/view/2126

[8] A. Sai Kumar, U. Siddhesh, N. Sai kiran and K. Bhavitha, "Design of High Speed 8-bit Vedic Multiplier using Brent Kung Adders," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-5, doi: 10.1109/ICCCNT54827.2022.9984591.

[9] J. Deng, A. Lin, G. Wan, Stanford University CNFET Model, Stanford University Nanoelectronics Lab, accessed on 22 January 2023, <https://nano.stanford.edu/downloads/stanford-cnfet-model>

[10] R. Martel, H. . -S. P. Wong, K. Chan and P. Avouris, "Carbon nanotube field effect transistors for logic applications," International Electron Devices Meeting. Technical Digest (Cat. No.01CH37224), Washington, DC, USA, 2001, pp. 7.5.1-7.5.4, doi: 10.1109/IEDM.2001.979456.

[11] R. Sahoo, R.R. Mishra, "Simulations of Carbon Nanotube Field Effect Transistors", International Journal of Electronic Engineering Research, ISSN 0975- 6450 Volume 1 Number 2 (2009) pp. 117–125

[12] Javey, A., Guo, J., Wang, Q. et al. "Ballistic carbon nanotube field-effect transistors", Nature 424, 654–657 (2003). https://doi.org/10.1038/nature01797

[13] Y. Ohno et. al., "Chirality assignment of individual single-walled carbon nanotubes in Carbon nanotube field-effect transistors by micro-photocurrent spectroscopy," Appl. Phys. Lett., vol. 84, no. 8, pp. 1368–1370, Feb. 2004.

[14] Dresselhaus, M.; Dresselhaus, G.; Saito, Riichiro, "Carbon fibers based on C60 and their symmetry" Physical Review B ,Volume 45, number 11, 1992

[15] J. Deng and H. . -S. P. Wong, "A Compact SPICE Model for Carbon-Nanotube Field-Effect Transistors Including Nonidealities and Its Application—Part I: Model of the Intrinsic Channel Region," in IEEE Transactions on Electron Devices, vol. 54, no. 12, pp. 3186-3194, Dec. 2007, doi: 10.1109/TED.2007.909030.

[16] J. Deng and H. . -S. P. Wong, "A Compact SPICE Model for Carbon-Nanotube Field-Effect Transistors Including Nonidealities and Its Application—Part II: Full Device Model and Circuit Performance Benchmarking," in IEEE Transactions on Electron Devices, vol. 54, no. 12, pp. 3195-3205, Dec. 2007, doi: 10.1109/TED.2007.909043.

[17] S. K. Sinha, P. Singh and S. Chaudhury, "Effect of temperature and chiral vector on emerging CNTFET device," 2014 International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2014, pp. 432-435, doi: 10.1109/IndiaCom.2014.6828174.

[18] A. Maiti, A. Svizhenko, and M. P. Anantram, 'Electronic Transport through Carbon Nanotubes: Effects of Structural Deformation and Tube Chirality', Phys. Rev. Lett., vol. 88, p. 126805, Mar. 2002.

[19] P. Prakash, K. Mohana Sundaram, M. Anto Bennet, "A review on carbon nanotube field effect transistors (CNTFETs) for ultra-low power applications", Renewable and Sustainable Energy Reviews, Volume 89, 2018, Pages 194-203, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2018.03.021.

[20] Avouris, P., Freitag, M. & Perebeinos, V. Carbon-nanotube photonics and optoelectronics. Nature Photon 2, 341–350 (2008). https://doi.org/10.1038/nphoton.2008.94

[21] B N, Shobha & Muniraj, Njr. (2014). Modeling and simulation of cntfet for prostate cancer detection. Journal of Theoretical and Applied Information Technology. 69. 530-538.

[22] Kerman, K., Morita, Y., Takamura, Y. et al. Peptide nucleic acid-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:40:09 UTC from IEEE Xplore. Restrictions apply.

modified carbon nanotube field-effect transistor for ultra-sensitive real-time detection of DNA hybridization. Nanobiotechnol 1, 65–70 (2005). https://doi.org/10.1385/NBT:1:1:065

[23] Oh Jeseung, et. al., "A carbon nanotube metal semiconductor field effect transistor-based biosensor for detection of amyloid-beta in human serum". Biosens Bioelectron 2013;50:345–50.

[24] Yang, L., Wang, S., Zeng, Q., Zhang, Z. and Peng, L.-M. (2013), Carbon Nanotube Photoelectronic and Photovoltaic Devices and their Applications in Infrared Detection. Small, 9: 1225-1236. https://doi.org/10.1002/smll.201203151

[25] "Vedic mathematics", Swami Sri Bharati Krsna Thirthaji Maharaja, Motilal Banarasidass Indological publishers and Book sellers, 1965.

[26] Jitendra Kumar Saini, Avireni Srinivasulu, Renu Kumawat, Fast and energy efficient full adder circuit using 14 CNFETs, Solid State Electronics Letters, Volume 2, 2020, Pages 67-78, ISSN 2589-2088, https://doi.org/10.1016/j.ssel.2020.09.002.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:40:09 UTC from IEEE Xplore. Restrictions apply.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Detection and Classification of Diabetic and Hypertensive Retinopathy Using CNN & Autoencoder

Amardeep Singh Kapoor
Department of Information Technology
Delhi Technological University
Delhi, India
amardeep08112000@gmail.com

Akshat Jain
Department of Information Technology
Delhi Technological University
Delhi, India
akshat87001@gmail.com

Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological University
Delhi, India
dinesh@dtu.ac.in

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205818

***Abstract*— In today's modern era where the world has advanced with latest technologies, still there is scope of improvement in various fields. Currently we are witnessing modern techniques using Artificial Intelligence and Machine learning. We can use them to enhance and improve the human health and lifestyle. One research area less explored is Diabetic Retinopathy and Hypertensive Retinopathy. Both are caused by Chronic Diabetic conditions or Hypertension Condition, which can cause serious damages in one's eye including blood vessel rupture cotton wool-spots etc. We can capture high resolution images and can use them to classify and detect the condition and help the patient with early diagnosis that can save them from losing their eyesight.**

***Keywords*— Deep Learning, Retina, Autoencoder, Hypertension Retinopathy, Diabetic Retinopathy, Hard Exudates, Cotton Wool**

I. INTRODUCTION

The eye consists of many membranes, one of which include the inner layer called the retina, which lies in the choroid and receives almost all nourishment from the blood vessels inside the choroid, with the rest coming from the retinal blood vessels that lie on the surface and can be seen with an ophthalmoscope. With Age retina tends to become weaker and more likely to be damaged. Retina is an important part and diagnosis of any disease in this part becomes a high priority. With tools like Machine Learning and AI we can help in preventing loss of eyesight in people. Usually there are two main diseases that are likely to cause harm to retina one is Diabetes and the other is Hypertension.

Hypertension - High blood pressure is a one of the conditions which his affecting the body's arteries. Also called Hypertension. In high blood pressure, the pressure of the blood through the arteries' walls is consistently very high. The heart would have to try harder and more to circulate blood. Blood pressure is measured in millimetres of mercury (mmHg). Hypertensive retinopathy is a condition of the blood vessels in the retina caused by high blood pressure. Symptoms usually occur in the later stages of the disease. On fundus examination, arteriovenous stenosis, arteriovenous dissection, changes in blood vessel walls, flaming or burning haemorrhage, cotton spots, yellow solid exudate, and papilledema were observed.

Diabetes - Diabetes is a disease that usually comes due to inefficient functioning of pancreas in creating insulin or the when body is not able to utilize use the insulin it produces. Insulin is a hormone that keep our blood sugar levels in check. Diabetes is of two types and usually people with diabetes mellitus are more prone to retinal disease called Diabetic retinopathy. Diabetic retinopathy is a condition caused due to advancement of diabetes. Diabetes can cause the tiny blood vessels in the retina to get swelled up which eventually led to bleeding and then leaking of fluid. In the eye, this process can lead to damage the retina. Usually, it affects our both eyes. Non-proliferative and proliferative are two main types of diabetic retinopathy. Non-proliferative: This is usually common type of retinopathy and occurs when the capillaries from backside of the eye get swelled up to form pockets. Non-proliferative retinopathy is categorized into three stages considering severity of the lesion. Proliferative: It is the advanced stage of diabetic eye condition. It occurs when the new blood vessels start growing in retina of eye.

II. RELATED WORK

From a long time, many studies have been done on automatic detection of diabetic retinopathy using different feature and techniques. M. Chetoui et al. [1] proposed a machine learning model to detect diabetic retinopathy from texture feature.U. Köse et al. [4] proposed a deep learning model that diagnosis diabetic retinopathy using convolutional neural networks.M. -C. Chang et al. [5] classified the severity of diabetic retiopathy using deep neural network. whereas talking about hypertensive retinopathy M. F. Syahputra et al. [6] proposed a

probabilistic neural network model to identify hypertensive retinopathy from retinal fundus images, V. Savant et al. [7] did analysis of the vessel parameters to help in detection of hypertensive retinopathy, N. Deepa et al. [11] that used a Dense Net on various features of hypertensive retinopathy to classify E-TLCNN, S. Irshad et al.[10] proposed a model to automatically detect cotton wool spots that helps in diagnosis of hypertensive retinopathy. There have been very few studies that actually classify diabetic and hypertensive retinopathy in an one single model with a good accuracy, because there are minor differences between them which can only be differentiated at an early stage of the disease by using the parameters of retinal arteries, veins and cotton wool spots.

III. AUTOENCODER

Autoencoder is generally used in the subject of unsupervised machine learning algorithms. You can use Encoders to compress the facts of images and reduce their dimensionality and through decoders, you can regenerate the original image. Autoencoders are generally compared with principal component analysis (PCA). PCA unearths the instructions along which you can undertake the records with the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:38 UTC from IEEE Xplore. Restrictions apply.

most variance, and in the case of Autoencoders reconstruct our original enter given only a compressed model of it. If anybody desires impressive statistics and wants to generate good results from their algorithms can reconstruct image data into encoded data and also can reconstruct the image data by use of autoencoders. An Autoencoders are a type of neural network that may learn how to regenerate photographs, textual content, and other features from compressed versions of themselves. An Autoencoder includes three layers: Encoding, Coding, and Decoding.

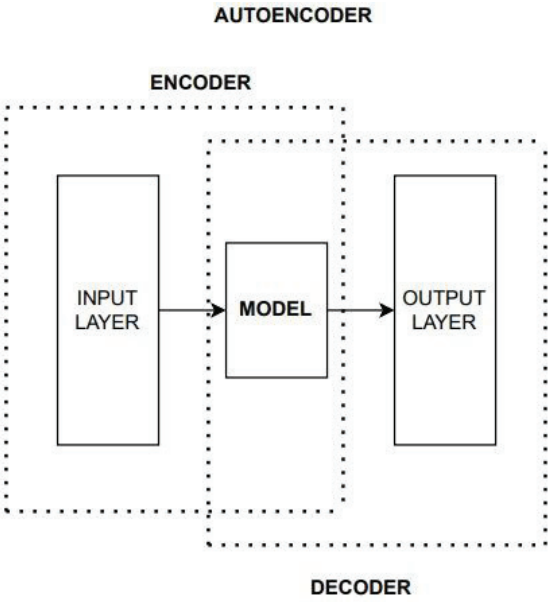


Fig. 1. Overview of Autoencoder

The Encoder layer is compressing the entered image into a compressed array which is in latent representation. It encodes the input photo as a compressed or encoded illustration in a discounted measurement. After processing of original image, the compressed image is distorted. In code, the layer consists of the user algorithm for encoding images. Compressed or input is passed to the decoder layer through code layer. The decoder layer decodes the encoded picture back to its original size. Latent space represents the decoded photo is reconstructed, and it ensures that there is no loss of data in the reconstructed image. The construction should be lossless construction.

IV. CNN MODEL

In this modern world deep learning techniques have proved to be a promising thing in medical field it helps in detecting and classifying diseases at an early stage. Deep Learning is a subfield of Machine learning that aims to learn from the data or images in a unsupervised way also called unsupervised learning. CNN or Convolutional Neural Network consist of several layers that include pooling , convolutional, fully connected and SoftMax layer. The output of each layer is passed through an activation function. At each convolutional layer a set of kernel is used to extract features depending on the depth of layer. Pooling layer comes into use when we have to reduce the complexity and or number of parameters. At the end comes the SoftMax layer for calculating the cross entropy loss and this loss is reduced by using stochastic gradient descent, which is achieved by readjusting weights and biases.

V. METHODOLOGY

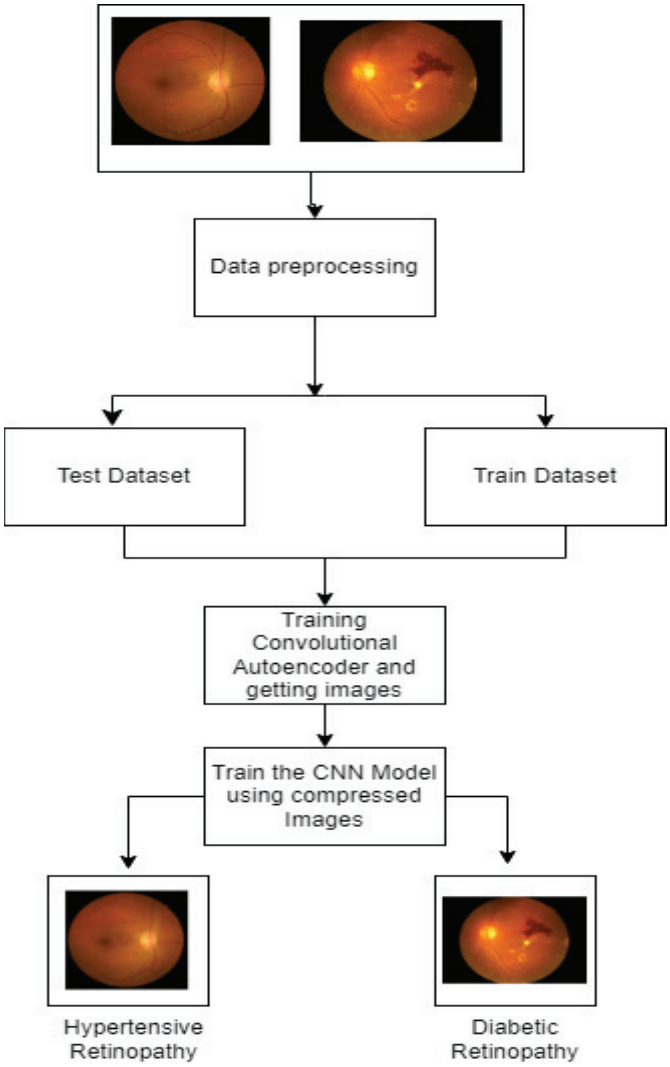


Fig. 2. Overview of Methodology

A. Dataset

The Dataset was collected from all sources available on internet some major sources for diabetic retinopathy specific fundus images were Kaggle which contains around 2100 images, MESSIDOR-2 Dataset that contains 1748 images, IDRiD Dataset that contains around 500 images.

Diabetic retinopathy Dataset contained different labels but we only needed a single label so we cleared the labels and classified the data under one label named Diabetic Retinopathy. For Hypertensive Retinopathy opensource dataset was not available so we collected the data from local eyecare centres and hospitals which was further pre-processed and labelled manually by experts as Hypertensive Retinopathy.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:38 UTC from IEEE Xplore. Restrictions apply.

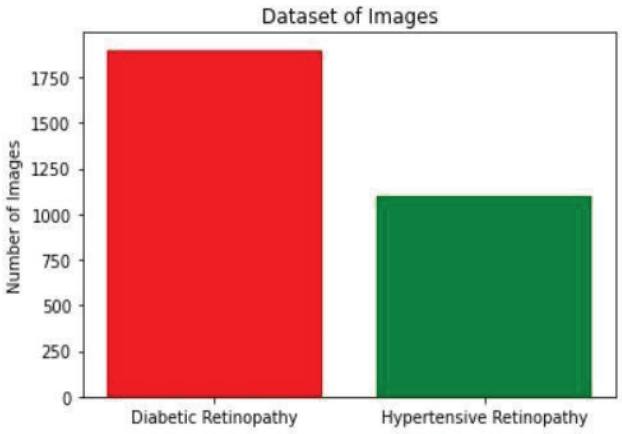


Fig. 3. Number of Images of Diabetic and Hypertensive Retinopathy

B. *Pre-processing*

The original dataset had been collected from various sources. It was in many formats like PNG, JPG, etc. It was converted to JPG format for better clarity and understanding of data and it also reduce the space with the help of the pillow library in python. The next step was the augmentation of images. It is a technique by which we can reduce or expand the size of the dataset. It is very helpful if our dataset is too large or too small.

As we collected data from various sources therefore the image size varies a lot. For tackling this we resized all the images into 256*256 matrix. Then the next step of pre-processing was the normalization of images.

For Normalization, we included Gaussian Blur Technique. It consists of two main parameters i.e., sigma and kernel size. Sigma is equivalent to the radius of the retina. Both the parameters were set to (0,0) for better clarity of the image. Then we pass data into encoders. The last part of pre-processing was to eliminate the excess part of the image which had no impact.

C. *Feature Extraction*

Retinal arteries and vein's width is calculated which makes up the Arterial Venous (AV) ratio. AVR is calculated as the ratio of the mean arteriolar diameter (AD) to the mean venous diameter (VD) of the vessels within the measurement area. There are two methods to calculate the AVR- it is calculated using the average of AW (Articole Width) and VW (Venules Width), the sum of the AW and VW, sum of the square of the AW and VW, CRAT (Central Retinal Artery Equivalent), and CRVA (Central Retinal Venous Equivalent). These values are repeated and hence the AVR is determined.

The other method to determine AVR is by classifying retinal vessels based on colour variations. The condition is known as copper wiring and if the conditions get worse it is known as silver wiring. As we know that there are some steps implemented for the detection of Hypertensive and Diabetic Retinopathy that use colour information. Therefore, it is important to normalize the image before the extraction of features. We already normalize the images by using encoders and gaussian normalization.

Using histogram stretched processed images helps to achieve normalized entropy for all available images, this is batter extraction technique. After this, colour normalization is performed by using the mean intensity value as 100, 75 and 40 for red, green, and blue respectively. Accurate blood vessel segmentation is an important step that helps us to achieve vessel measurements such as width for Arterial Venous ratio.

The hard exudates are composed of lipid and proteinaceous material. It is generally found in Hypertensive Retinopathy. These exudates are capable of building a star shape within the macula, especially when hypertension is critical. In serious hypertensive Retinopathy, the optic disk becomes choked and swollen with excessive accumulation of fluid. It is very helpful for classifying the images as hypertensive Retinopathy.

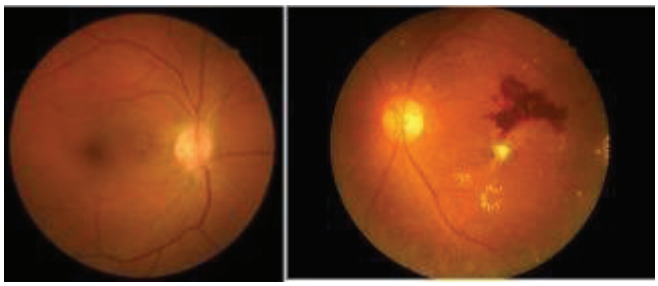


Fig. 4. Images showing exudates, cotton wool, Haemorrhage in Diabetic and Hypertensive Retinopathy.

VI. MODEL BUILDING

To carry out this research we have built a Convolutional Neural Network (CNN) model. This model will accurately classify the fundus images into Diabetic or Hypertensive Retinopathy 2D Convolutional, Max Pooling, Fully Connected, and SoftMax layers makes this proposed CNN Model.

For the first stage encoded RGB Fundus images are resized into 512*512*3 for the input layer of CNN that was reduced in dimensionality using autoencoders. Then comes a 2D convolutional input layer with 5 kernels and 1 stride taking each filter size to 5*5*3. Then comes the max pooling layer with 5 filter and 2 stride that reduces the feature map to 254*254 by doing all this we have reduced the high dimension images to low dimension with maintaining the quality by using autoencoders, this all thing would help in increasing the processing speed of Convolutional Neural Network. After this a fully connected layer is deployed output which corresponds to 2 different classes either diabetic or hypertensive retinopathy Finally the actual probability of each output class is calculated using SoftMax function.

VII. RESULT AND ANALYSIS

In this paper, we were using dataset of 3000 images in which 1100 was images of Hypertensive Retinopathy and remaining was images of Diabetic Retinopathy. Rather than training the whole set of images. The images were split into batches of 100 images and we have a 30 batches of images. In the current study, we trained the network for 25 and 30 epochs where a single epoch represents the 30 iterations. Leave out one batch for cross validation. Repeated this steps 30 times. We were using the learning rate around 10^ (-5). Training and testing accuracies were computed using the following equations.

$$Accuracy = \frac{TP+FN}{TP+FP+TN+FN} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:38 UTC from IEEE Xplore. Restrictions apply.

The graph of training accuracy and validation accuracy show in fig 5. Loss and validation loss were computed using the following equations.

$$A_{loss_train} = \frac{\sum_i^{31}\sum_j^{30} z(i,j)}{31} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

$$A_{loss_test} = \frac{\sum_i^{31}\sum_j^{30} z(i,j)}{31} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

z(i,j) was a function which was calculated the root mean square difference between actual and predicted values. The graph of training loss and validation loss accuracy and validation accuracy show in fig 6.

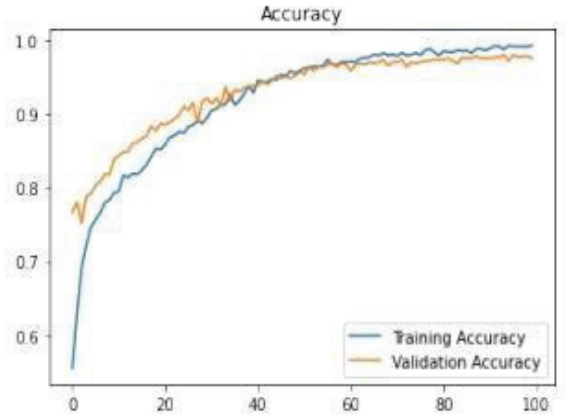


Fig. 5. Accuracy graph

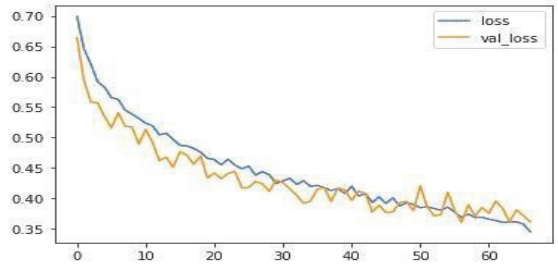


Fig. 6. Loss During Training

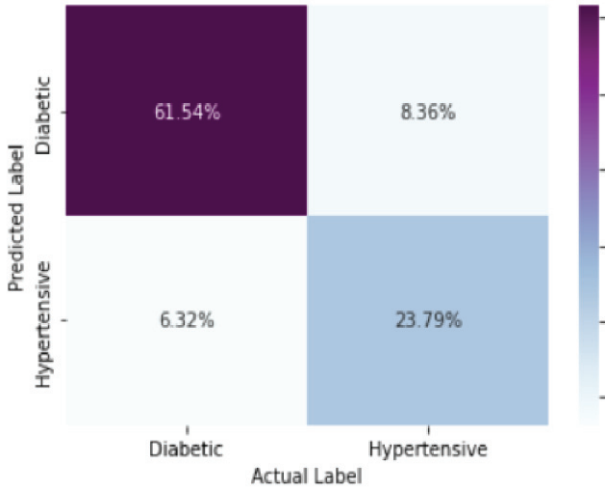


Fig. 7. Confusion Matrix of the CNN Model.

The fig 7 showing the confusion matrix of CNN model for Diabetic and Hypertensive Retinopathy. It shows percentage of Diabetic Retinopathy and Hypertensive Retinopathy Predicting correctly and incorrectly. For calculating the accuracies, we need to add the percentages of images predicted correctly. Let suppose we have two types of images then the formula of Accuracy is Percentage of Images of type1 predicted correctly and percentage of images of type 2 predicted correctly.

The accuracy after all the training and testing was around 85.33% which was tested on 1100 images for hypertension and 1900 images for diabetes.

VIII. CONCLUSIONS

As it has been shown above, we have proposed an efficient and highly accurate framework to detect and classify Diabetic retinopathy and hypertensive retinopathy by using Fundus Photography. We have designed the algorithm that can use different features like exudates, haemorrhage, cotton wool and many more to detect the disease and differentiates between Diabetic retinopathy and Hypertensive retinopathy, which helps the patient to diagnose and treat the disease efficiently and quickly.

REFERENCES

[1] M. Chetoui, M. A. Akhloufi and M. Kardouchi, "Diabetic Retinopathy Detection Using Machine Learning and Texture Features," 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE), Quebec, QC, Canada, 2018, pp. 1-4, doi: 10.1109/CCECE.2018.8447809..

[2] H. Jiang, K. Yang, M. Gao, D. Zhang, H. Ma and W. Qian, "An Interpretable Ensemble Deep Learning Model for Diabetic Retinopathy Disease Classification," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 2045-2048, doi: 10.1109/EMBC.2019.8857160.K. O'Shea and R. Nash, An Introduction to Convolutional Neural Networks, Neural and Evolutionary Computing, Cornell University Library, 2016..

[3] V. Vipparthi, D. R. Rao, S. Mullu and V. Patlolla, "Diabetic Retinopathy Classification using Deep Learning Techniques," 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2022, pp. 840-846, doi: 10.1109/ICESC54411.2022.9885687.

[4] Ö. Deperlıoğlu and U. Köse, "Diagnosis of Diabetic Retinopathy by Using Image Processing and Convolutional Neural Network," 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 2018, pp. 1-5, doi: 10.1109/ISMSIT.2018.8567055.

[5] C. -Y. Chen and M. -C. Chang, "Using Deep Neural Networks to Classify the Severity of Diabetic Retinopathy," 2022 IEEE International Conference on Consumer Electronics - Taiwan, Taipei, Taiwan, 2022, pp. 241-242, doi: 10.1109/ICCE-Taiwan55306.2022.9869142.

[6] M. F. Syahputra, Nurrahmadayeni, I. Aulia and R. F. Rahmat, "Hypertensive retinopathy identification from retinal fundus image using probabilistic neural network," 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), Denpasar, Indonesia, 2017, pp. 1-6, doi: 10.1109/ICAICTA.2017.8090989.

[7] V. Savant and N. Shenvi, "Analysis of the Vessel Parameters for the Detection of Hypertensive Retinopathy," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2019, pp. 838-841, doi: 10.1109/ICECA.2019.8821850.

[8] C. Agurto, V. Joshi, S. Nemeth, P. Soliz and S. Barriga, "Detection of hypertensive retinopathy using vessel measurements and textural features," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 2014, pp. 5406-5409, doi: 10.1109/EMBC.2014.6944848.

[9] S. Irshad and M. U. Akram, "Classification of retinal vessels into arteries and veins for detection of hypertensive retinopathy," 2014 Cairo International Biomedical Engineering Conference (CIBEC), Giza, Egypt, 2014, pp. 133-136, doi: 10.1109/CIBEC.2014.7020937.

[10] S. Irshad, M. Salman, M. U. Akram and U. Yasin, "Automated detection of Cotton Wool Spots for the diagnosis of Hypertensive Retinopathy,"

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:38 UTC from IEEE Xplore. Restrictions apply.

2014 Cairo International Biomedical Engineering Conference (CIBEC), Giza, Egypt, 2014, pp. 121-124, doi: 10.1109/CIBEC.2014.7020932.

[11] N. Deepa and D. T, "E-TLCNN Classification using DenseNet on Various Features of Hypertensive Retinopathy (HR) for Predicting the Accuracy," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 1648-1652, doi: 10.1109/ICICCS51141.2021.9432255.

[12] A. Montoro, S. Morales, V. Naranjo, F. López-Mir and M. Alcañiz, "Feature extraction for retinal vascular network classification," IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain, 2014, pp. 404-407, doi: 10.1109/BHI.2014.6864388.

[13] M. S. Ahmed, T. Mithun, R. Nagaraju, Y. Sowmya Reddy, W. Madhukar Baban and S. Jaiswal, "Integration of Artificial Intelligence for a low-cost diagnosis of Diabetic Retinopathy," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2021, pp. 1750-1753, doi: 10.1109/ICECA52323.2021.9676109.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:52:38 UTC from IEEE Xplore. Restrictions apply.

European Polymer Journal 197 (2023) 112340

Contents lists available at ScienceDirect

European Polymer Journal

journal homepage: www.elsevier.com/locate/europolj

Development and assessment of carboxymethyl tamarind kernel gum-based pH-responsive hydrogel for release of diclofenac sodium

Priyanka Meena , Poonam Singh *, Sudhir G. Warkar *

Department of Applied Chemistry, Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
Smart hydrogel
CMTKG
Diclofenac sodium
Biopolymer
Sol-gel

ABSTRACT

The current research work focuses on the fabrication of carboxymethyl tamarind kernel gum/polyvinylpyrrolidone/polyacrylamide (CMTKG/PVP/PAM) based hydrogel and its loading was done with diclofenac sodium drug (DS) to obtain smart pH-responsive hydrogels. In addition, the effect of concentration of cross-linker- N, N′-methylene bis(acrylamide) (MBA), biopolymer (CMTKG), as well as initiator- potassium persulphate (KPS) on the swelling behavior of smart hydrogel was also analyzed. The synthesized hydrogels were characterized using Scanning Electron Microscopy (SEM), Powder X-ray Diffraction (PXRD), Thermogravimetry Analysis (TGA), and Attenuated Total Reflection-Fourier Transform Infrared spectroscopy (ATR-FTIR). ATR-FTIR confirms the successful formation of hydrogels, while PXRD and SEM indicate the amorphous and porous nature of the polymeric network. The TGA results indicate higher thermal stability for DS-loaded hydrogel than unloaded hydrogel. However, sol–gel analysis shows a rise in gel fraction, with increased concentrations of CMTKG, MBA, and KPS. The swelling ratio as well as the release of DS were found to be higher under slightly alkaline conditions. The result of drug loading and release indicates that PVP increases drug loading (%) and release (%). The drug release behavior of DS-loaded hydrogel follows the Fickian mechanism of diffusion with the Korsmeyer-Peppas model as the best-fitted one. The synthesized DS-loaded CMTKG/PVP/PAM hydrogels could be a promising substitute for prolonged delivery of sparingly soluble DS in pH 7.4.

1. Introduction

A drug delivery system (DDS) acts as a sustained release system that can alter the pharmacokinetics and biodistribution of the drug by enhancing several crucial properties of drugs i.e. solubility, in vivo stability, specificity, efficacy, etc [1]. Till today numerous materials have been explored in the drug delivery field, including nano-scale materials (nano-capsules, dendrimers, nanotubes, nano gels, and so on) [2,3], and inorganic-based DDS (layer double hydroxide, graphene oxide, metal oxides etc) [1,4,5]. However, conventional drug delivery systems do possess certain limitations i.e. poor bioavailability, low biocompatibility, high toxicity, burst drug release, and lesser hydrophilicity, resulting in reduced drug delivery efficiency. To address the above-mentioned issues, drug delivery systems based on hydrogels have been focused in recent decades owing to their easily tunable properties during synthesis, followed by controlled release, degradation, and efficiency to protect labile drugs [6].

Basically, hydrogels are highly hydrated 3D networks, derived from natural, semi-synthetic, or synthetic polymers having physical or covalent cross-linkage [7,8]. One of the most remarkable properties of hydrogel to be termed as smart or stimuli-sensitive hydrogels is their ability to adjust their volume in response to changes in physical and chemical stimuli such as pH, light, magnetic field, temperature, enzyme, ionic strength, etc [9,10,11,12]. The most notable among all are pH-responsive hydrogels, which provide an excellent drug administration system [11].

Polyacrylamide (PAM) is one of the most extensively used pH-sensitive polymers in drug delivery applications [13]. It is a synthetic, non-toxic polymer with excellent mechanical strength, and contains a cationic amine functional group that either protonates or deprotonates to show variation in swelling with a change in pH [14,15]. PAM has been widely combined with natural polymers such as gelatin, carboxymethyl tamarind kernel gum, collagen etc. to synthesize composite hydrogels [16,17]. The introduction of biopolymers imparts efficient drug release and also enhances the hydrophilicity, biodegradability, and biocompatibility of the hydrogel [11,18,19].

Tamarind kernel gum (TKG) is a biopolymer obtained from Tamarindus indica L., a typical African tree and carboxymethylation of TKG

* Corresponding authors.
E-mail addresses: poonam@dtu.ac.in (P. Singh), sgwarkar@dtu.ac.in (S.G. Warkar).

https://doi.org/10.1016/j.eurpolymj.2023.112340
Received 30 April 2023; Received in revised form 27 July 2023; Accepted 30 July 2023
Available online 31 July 2023
0014-3057/© 2023 Elsevier Ltd. All rights reserved.

results in the formation of carboxymethyl tamarind kernel gum (CMTKG) [20]. CMTKG possesses a high swelling capacity, high viscosity, and enhanced hydrophilicity compared to TKG [21]. It is a water-soluble polymer which contains glucose, xylose, and galactose in the molar ratio of 3:2:1 [22,23]. With variations in the pH of the medium, COO^- groups in this polymer undergo deprotonation or protonation, resulting in pH-dependent drug delivery [24].

Owing to the hydrophilic nature of hydrogels, they have been most commonly utilized for delivering hydrophilic drugs such as moxifloxacin hydrochloride, acyclovir [25,26,27]. To increase the compatibility of hydrophobic drugs with the polymer matrix, either the polymers capable of the formation of inclusion complexes such as cyclodextrin are introduced or various hydrophobic components are incorporated into the hydrogels. Polyvinylpyrrolidone, poly(lactic acid), nanoparticles, etc, are some of the hydrophobic components incorporated into the polymeric matrix [25]. Polyvinylpyrrolidone (PVP) is a non-toxic, and biocompatible polymer mainly used for drug delivery applications. It is an amphiphilic polymer containing a pyrrolidone ring, C = O, C-N, and CH_2 as a hydrophilic component and an alkyl group as a hydrophobic part [28,29]. The hydrophobic component of PVP interacts with the hydrophobic drug and increases drug loading capacity. Therefore, PVP is employed to deliver hydrophilic, hydrophobic as well as sparingly soluble drugs such as amoxicillin [30], acyclovir [31], diclofenac sodium [32], etc.

Diclofenac sodium (DS) is one of the most commonly prescribed nonsteroidal anti-inflammatory drugs (NSAIDs). It is sodium salt of o-(2,6-dichlorophenylamino)-phenylacetic acid and its antipyretic, analgesic, and anti-inflammatory actions help to treat rheumatoid arthritis and osteoarthritis. The DS half-life is relatively less (1–2 h) and so rapidly eliminated from the body. As a result, numerous doses of DS are required to provide sustained pharmacological action which causes unfavorable side effects such as ulcers and gastrointestinal bleeding. Therefore, a novel delivery system for DS release is required to decrease the complications and to ensure the prolonged delivery of DS [33,34].

In the present research work, first CMTKG/PVP/PAM-based hydrogel was synthesized and was then loaded using DS to obtain smart pH-responsive hydrogels for prolonged release of DS. Incorporation of PVP was done so as to improve the loading capacity of sparingly soluble DS drugs. The synthesized hydrogel was characterized using ATR-FTIR, SEM, PXRD, TGA, and to confirm the formation, surface morphology, crystallinity, and thermal stability of the developed hydrogel. A sol–gel study was conducted to evaluate the hydrogel's sol and gel portions. Finally, the impact of amounts of CMTKG, KPS, or MBA on swelling ratio was also examined for all DS-loaded hydrogels using pH 1.2 and 7.4 buffer. The work also emphasizes upon *in vitro* release of DS to estimate the DS release percentage. To contemplate the drug release mechanism, various kinetic models like Zero-Order, Hixson- Crowell, Higuchi, Korsmeyer-Peppas, and First-Order were employed.

2. Materials and methods

2.1. Materials

Acrylamide (CDH, New Delhi), Polyvinylpyrrolidone K-30 (CDH, New Delhi), Potassium persulphate (Fischer Scientific, Mumbai), and N, N′-methylenebis(acrylamide) (Merck, Germany), were used as supplied. Carboxymethyl Tamarind Kernel Gum was gratefully provided by Hindustan Gum and Chemicals Ltd., Haryana, India and Diclofenac sodium was provided by Unicure India Pvt Ltd, Noida, India.

2.2. Synthesis of CMTKG/PVP/PAM hydrogel and DS-loaded hydrogels

The novel hydrogel based on CMTKG, PVP, and AM was synthesized using a free radical mechanism by employing KPS (initiator) and MBA (cross-linker), shown in Fig. 1. The reaction was typically carried out by dissolving the desired quantity of CMTKG, AM, and PVP in distilled water on a magnetic stirrer for 1hr. To the resultant solution, KPS and MBA were added and stirred. Later, the solution was transferred into 20 mL test tubes and placed in a water bath at 60 °C for 1 h. The obtained hydrogel was cut into discs, dried at room temperature, and then in the oven at 60 °C till a persistent weight was attained [35,36]. For the synthesis of DS-loaded hydrogels, during the addition of KPS, and MBA, 50 mg diclofenac sodium dissolved in 1 mL ethanol was added to the mixture, however, the rest procedure followed the same as above. Table 1 summarizes the exact ratios of the KPS, MBA, CMTKG, PVP, and AM used during the synthesis of hydrogels [37].

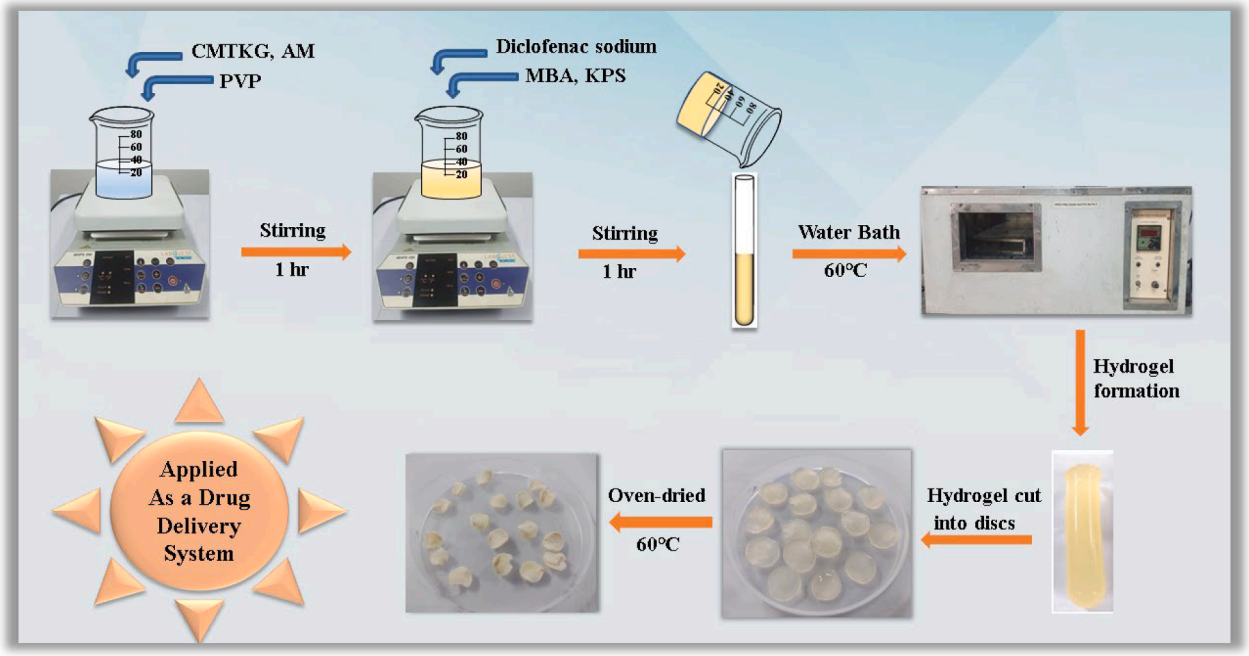


Fig. 1. Pictorial representation of the fabrication of DS-loaded CMTKG/PVP/PAM hydrogel.

Table 1
Different ratios of reagents required in the synthesis of the hydrogels and their swelling ratio.

Sample Code	CMTKG (g)	AM (g)	PVP (g)	MBA (mg)	KPS (mg)	DS (mg)	Swelling Ratio (%)		Gel fraction (%)
							pH 1.2	pH 7.4	
A-1	0.20	2.5	0.1	17	35	50	864	1035	73
A-2	0.25	2.5	0.1	17	35	50	886	1062	75
A-3	0.30	2.5	0.1	17	35	50	915	1081	76
A-4	0.35	2.5	0.1	17	35	50	941	1100	78
A-5	**0.40**	**2.5**	**0.1**	**17**	**35**	**50**	**967**	**1128**	80
A-6	0.40	2.5	0.1	17	45	50	879	1042	77
A-7	0.40	2.5	0.1	17	40	50	939	1088	78
A-8	0.40	2.5	0.1	17	30	50	919	1052	79
A-9	0.40	2.5	0.1	17	25	50	799	970	76
A-10	0.40	2.5	0.1	27	35	50	819	1024	81
A-11	0.40	2.5	0.1	37	35	50	714	916	82
A-12	0.40	2.5	0.1	47	35	50	642	845	84
A-13	0.40	2.5	0.1	57	35	50	533	737	85
A-14	0.40	2.5	00	17	35	50	960	1123	77
A-15	0.40	2.5	0.1	17	35	0	949	1109	82

2.3. Swelling studies

Gravimetric analysis was carried out to analyse the swelling nature of the prepared DS-loaded hydrogels. First, dried hydrogels were weighed and soaked in pH 1.2 and 7.4 buffer solutions. After a regular interval, the swollen hydrogel disc was taken out, surface water was wiped using filter paper, and again, weight was noted [33]. The study was carried out in triplicate, and the swelling ratio was computed using the given equation.

$$\text{Swelling Ratio}\,(\%) = \frac{W_S - W_D}{W_D} \times 100 \quad (1)$$

Here, W_S refers to the weight of the swollen hydrogel at time t, whereas W_D refers to the initial weight of the dried hydrogel [38].

2.4. Sol-Gel analysis

Sol-gel analysis was done to discover the uncrosslinked portion of the synthesized DS-loaded hydrogel. A gel is the insoluble cross-linked portion of a hydrogel, and a sol implies the soluble uncross-linked portion of a hydrogel [33]. For this analysis, the dried hydrogel disc was weighed and processed to the soxhlet extraction method using distilled water as solvent at 100℃ for a duration of 4 hr. After 4 hr, the extracted hydrogel disc was dried at 45 °C in an oven till persistent weight was attained. The sol and gel fraction calculation was performed using the following equations.

$$\text{Sol Fraction} = \frac{W_i - W_F}{W_F} \times 100 \quad (2)$$

$$\text{Gel fraction} = 100 - \text{Sol fraction} \quad (3)$$

Here, W_i implies the weight of the initially dried hydrogel, and W_F implies the final weight of extracted dried hydrogel after being subjected to soxhlet extraction [39].

2.5. Drug loading and entrapment efficiency

The ideal candidate, hydrogel A-5, was selected for drug loading and entrapment studies on account of its highest swelling ratio of about 1128% and 967% at pH 7.4 and 1.2 respectively. The effect of PVP (amphiphilic polymer) on drug loading & entrapment efficiency was established by comparing formulation A-5 (CMTKG/PVP/PAM) that contains PVP to formulation A-14 (CMTKG/PAM) hydrogel, which lacks PVP. The drug loading (DL %) & entrapment efficiency (DEE %) was determined by placing the pre-weighed hydrogels (0.1 g) into a physiological buffer of pH 7.4 for 24 hr. The solution was withdrawn after 24 hr, and absorbance was noted at 276 nm with a UV–visible spectrophotometer (Model: Cary 300 UV–Vis). The amount of DS was estimated using a calibration curve. The below equations were used to establish the percent of drug loading and drug entrapment efficiency [40].

$$\text{DL}\,(\%) = \frac{\text{Drug amount in hydrogel}}{\text{Weight of hydrogel disc}} \times 100 \quad (4)$$

$$\text{DEE}\,(\%) = \frac{\text{Drug amount in hydrogel}}{\text{Theoretical Drug amount in hydrogel disc}} \times 100 \quad (5)$$

2.6. In vitro release study of diclofenac sodium

In vitro release analysis of DS was accomplished at physiological pH of 7.4 and 1.2 using DS-loaded CMTKG/PVP/PAM (A-5) and DS-loaded CMTKG/PAM (A-14) hydrogel to study the impact of the incorporation of PVP on drug release. The release study was done in an orbital incubator shaker by immersing 0.1 g of the DS-loaded hydrogel disc into 100 mL buffer of pH 1.2 and 7.4 at 37 °C, respectively [40]. After 1 hr, a 3 mL sample was extracted, and an equal amount of fresh buffer was made-up. Absorbance was recorded on a UV–visible spectrophotometer at λ_{max} 276 nm. The study was performed in triplicate and the drug content was evaluated using a calibration curve [41].

2.7. Release kinetics studies of diclofenac sodium

Several Mathematical models, such as the Zero-Order, Hixson- Crowell, Higuchi, Korsmeyer-Peppas, and First-Order, were used to identify the release mechanism of the drug. The release data were fitted to all the above-mentioned models, and regression coefficient (R^2) values were compared. The R^2 value must be closer to 1 for the model to be considered the best-fit model [42].

2.8. Characterization

PXRD was recorded using a Bruker D8 diffractometer within 2θ range of 5-70℃ using Cu Kα radiation. The ATR-FTIR analysis was done by an ATR-FTIR spectrophotometer (Model: PerkinElmer spectrum 2). The TGA was done by PerkinElmer TGA 4000 with an N_2 atmosphere in the range of 25 °C–900 °C at a rate of 10 °C/min. The surface morphology was investigated using SEM (Model: JEOL Japan Mode: JSM 6610LV).

3. Results and discussions

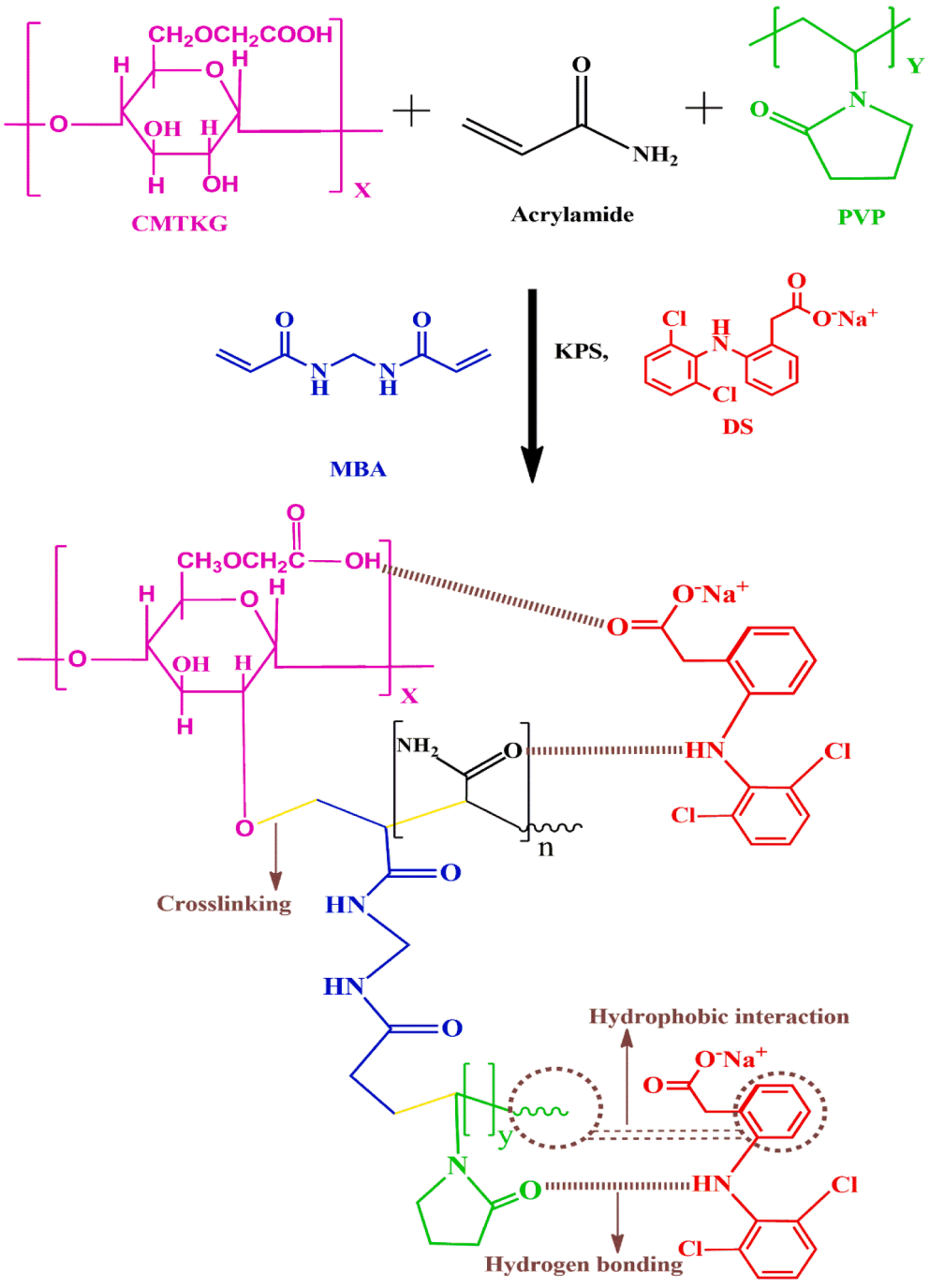
3.1. Mechanistic pathway followed for the formation of DS-loaded CMTKG/PVP/PAM hydrogel and their swelling studies

The details of various combinations (A-1 to A-13) of DS-loaded hydrogels attempted using a variable concentration of CMTKG, KPS, and MBA are given in Table 1. Here, KPS was used as an initiator, MBA was used as a cross-linker, and drug DS was loaded into the hydrogel. All the samples were fabricated with a varied amount of CMTKG (A1-A5), KPS(A6-A9), or MBA(A10-A13). The proposed mechanism involved during synthesis for hydrogel formation has been shown in Scheme 1. In the reaction, at first sulphate radical is formed by the decomposition of KPS (initiator) at 60 °C, which further leads to the polymerization of AM. While MBA acts as a cross-linker, the MBA helps in the formation of a crosslinked polymeric network based on CMTKG, PAM, and PVP. The hydrophobic interaction between the alkyl group of PVP and the aryl ring of DS, along with hydrogen bonding of DS with PVP, CMTKG and PAM, results in the synthesis of a tri-dimensional network of DS-loaded CMTKG/PVP/PAM hydrogel [43].

The swelling ratio of all DS-loaded hydrogels was examined using pH 1.2 and 7.4 buffer by varying amounts of CMTKG, KPS or MBA. The swelling ratio of all synthesized hydrogel were mentioned in Table 1.

3.1.1. Impact of biopolymer

The impact of concentrations of the CMTKG (biopolymer) on the swelling ratio, is presented in Fig. 2(a). The hydrogel swelling increases with an increase in the amount of CMTKG from 0.2 g to 0.4 g. Maximum



Scheme 1. Proposed mechanism for synthesis of DS-loaded CMTKG/PVP/PAM hydrogel.

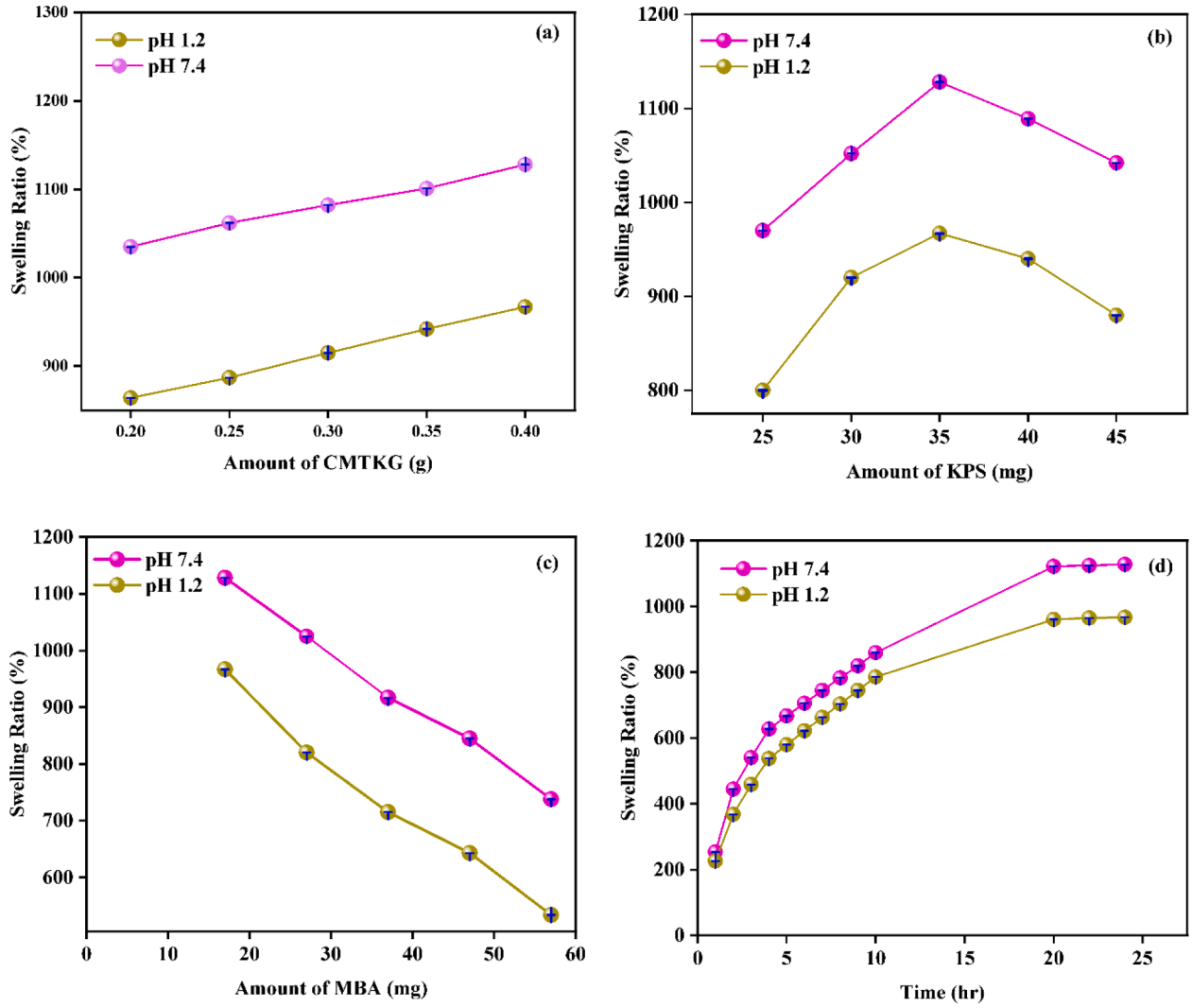


Fig. 2. Impact of the **(a)** CMTKG, **(b)** KPS, **(c)** MBA on swelling and **(d)** Rate of swelling of A-5 hydrogel in pH 1.2 and 7.4 with respect to time.

swelling of about 1128% was noticed at pH 7.4, while at pH 1.2, 967% swelling was recorded at 0.40 g of CMTKG. This may be associated with a greater number of carboxylate ions as the amount of CMTKG increases. The repulsion of carboxylate ion trigger chain relaxation and, as a consequence, swelling increases. However, attempts were made to process hydrogel with a concentration of CMTKG greater than 0.40 g but, the solution become highly viscous, and was challenging to stir [22].

3.1.2. *Impact of initiator*

Since the maximum swelling was observed at a biopolymer amount of 0.40 g, therefore it was kept constant. While the impact of altering the KPS concentration is shown in Fig. 2(b). As the KPS amount was elevated, the swelling percentage increased to a certain limit, and later increase in the KPS amount led to decreased water absorption capacity. At 35 mg of an initiator, maximum swelling of about 1128% was noticed in pH 7.4, while in pH 1.2, 967% swelling was reported. For the amount of KPS below 35 mg, the swelling ratio decreased to 723% at pH 7.4 and 620% at pH 1.2. This may be the consequence of a large quantity of unreacted soluble monomer, leading to a weak hydrogel network and reducing the swelling. When, the amount of KPS was increased beyond 35 mg, again swelling decreased because of the high collision between the free radicals. It leads to oligomer formation and increased cross-linking density; as a result, swelling decreases [22].

3.1.3. *Impact of cross-linker*

According to Flory's network theory, one of the significant factors that influence hydrogel swelling is cross-linking density. A slight variation in the cross-linker amount causes substantial changes in the swelling percentage, as presented in Fig. 2(c). Maximum swelling of about 1128% was noticed at pH 7.4. While at pH 1.2, maximum swelling of about 967% was reported at 17 mg of a crosslinker. When the amount of MBA is lesser than 17 mg, it was found that a jelly-like substance was produced due to a lack of a strong cross-linked network between polymeric chains [22]. The hydrogel swelling ratio decreased when the MBA concentration varied from 17 to 57 mg. A higher cross-linker amount reduces the free space inside the hydrogels and decreases the swelling of the hydrogels. Hence, the resulting rigid structure cannot expand, so water absorption decreases [34].

Among all the formulations attempted the greater swelling was noticed for A-5 hydrogel having CMTKG (0.40 g), AM (2.5 g), PVP (0.1 g), KPS (35 mg), MBA (17 mg), and DS (50 mg). However, for the A-5 sample, a higher swelling ratio was noticed in pH 7.4, than in pH 1.2, as shown in Fig. 2 (d). The greater swelling at alkaline pH is a consequence of the deprotonated carboxylate anion of CMTKG. This leads to anion-anion repulsion between COO^- ion, which further tends to polymeric chain relaxation and more free space resulting in fluid absorption [44]. While at pH 1.2, hydrogen bonding exists between PAM and CMTKG, causing shrinkage of the hydrogel network, thereby decreasing water absorption capacity [38].

3.2. Sol-Gel analysis

Sol-gel study for synthesized DS-loaded CMTKG/PVP/PAM hydrogel samples is shown in Fig. 3. An increase in gel fraction was noticed, with increases in the amount of CMTKG (A-1 A-2, A-3, A-4, A-5) and KPS (A-8, A-9). This is because more free radicals are generated and, therefore, there are more active sites for crosslinking, so the gel fraction increases. In addition, when KPS is further increased (A-6, A-7), a soluble oligomer chain forms, resulting in a greater sol fraction and a lesser gel fraction. As the concentration of MBA (A-10, A-11 A-12, A-13) increases, cross-linking density increases within the polymeric network, which causes an increase in gel fraction [33].

3.3. Characterization

PXRD was done to examine the crystallinity nature of hydrogel and the dispersion of the drug into the hydrogel matrix, as shown in Fig. 4. DS shows its characteristic peak at 2θ values of 6.2, 6.7, 8.6, 11.3, 15.2, 17.2, 20.3, 27.1, and 37.9 [43]. The sharp peaks in the DS reveal the crystallinity nature of the drug. However, the broad peak for CMTKG/PVP/PAM hydrogel indicated the hydrogel disc was amorphous and shows no long crystallographic order [38]. In the diffractogram of DS-loaded CMTKG/PAM/PVP hydrogel, no characteristic peak of DS was observed indicating uniform dispersion of DS in polymeric matrices [45], [46]. The increase in peak intensity was observed for DS-loaded hydrogel compared to unloaded hydrogel, which implies that the drug was effectively entrapped within the amorphous hydrogel matrices [33].

ATR-FTIR spectrum of CMTKG, AM, MBA, CMTKG/PVP/PAM, DS-loaded CMTKG/PVP/PAM (A-5) hydrogels and DS are presented in Fig. 5, and their ATR-FTIR data are presented in Table 2. In the FTIR spectra of AM, the peaks at 1665 cm^{-1} are assigned as -C = O asymmetric stretch, while in CMTKG/PVP/PAM and DS-loaded CMTKG/PVP/PAM hydrogel, moved to 1603 cm^{-1} and 1606 cm^{-1} respectively, validating the decrease in conjugation and confirmed the polymerization [35].

In the FTIR spectrum of MBA, the peaks at 1304 cm^{-1} correspond to C-N stretch, which in CMTKG/PVP/PAM and DS-loaded CMTKG/PVP/PAM hydrogel, moved to 1323 cm^{-1} and 1320 cm^{-1} respectively, indicate the lesser conjugation because of cross-linking. The peak at 1624 cm^{-1} represents the C = C stretch, which in CMTKG/PVP/PAM and DS-loaded CMTKG/PVP/PAM observed at 1659 cm^{-1} and 1660 cm^{-1}, confirming the cross-linking. This concludes that crosslinked CMTKG/PVP/PAM hydrogel has been successfully synthesized [43]. In

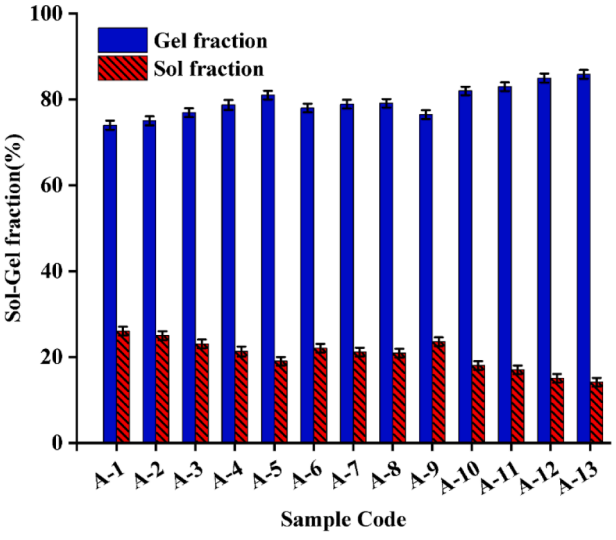


Fig. 3. Sol-Gel studies of synthesized formulations of DS-loaded CMTKG/PVP/PAM hydrogel.

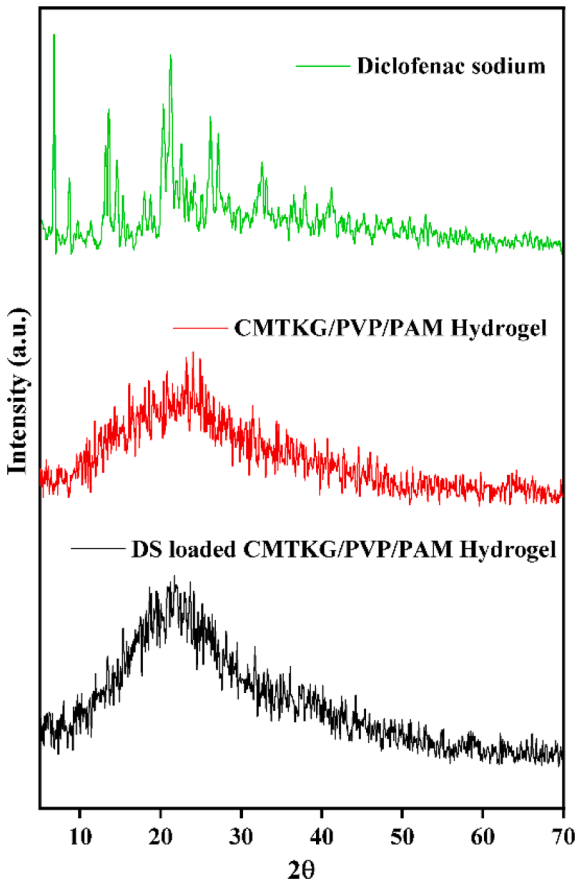


Fig. 4. PXRD pattern of DS, CMTKG/PVP/PAM hydrogel (A-15), DS-loaded CMTG/PVP/PAM hydrogel (A-5).

DS-loaded CMTKG/PVP/PAM, the new characteristic peak of DS is noticed at 2549 cm^{-1}, 2273 cm^{-1}, and 704 cm^{-1}. This signifies that drug is successfully encapsulated into the hydrogel matrix. There was no appearance or disappearance of additional peaks other than the drug noticed in DS-loaded hydrogel compared to unloaded hydrogel, which concludes that no chemical interaction exists, only physical interaction (hydrogen bonding) took place between the drug and polymeric network [38].

TGA was performed to analyse the thermal stability of hydrogel to understand the interaction existing between the drug and hydrogel, as presented in Fig. 6. In both CMTKG/PVP/PAM and DS-loaded CMTKG/PVP/PAM hydrogel (A-5), four step weight loss was noticed. The first stage is associated with the evaporation of H_2O molecule and initial decomposition of polymeric chain of CMTKG/PVP/PAM and DS-loaded CMTKG/PVP/PAM hydrogel, with a weight loss of 11.1% and 9.6% at 30–245 °C respectively. The second stage corresponds to the degradation of the polymer backbone, i.e. carboxymethyl and hydroxyl of CMTKG/PVP/PAM and DS-loaded CMTKG/PVP/PAM hydrogel, with the weight loss of 11% and 9.4% at 245–297 °C respectively. The third stage indicates the breakdown of crosslinking of the polymeric network, with weight loss of 31.8% and 31.1% at 297–385 °C. The fourth stage, corresponds to the complete degradation of the hydrogel, with weight loss of 40.2 % and 38.8% at 385–900 °C respectively [37].

The TGA thermogram of CMTKG/PVP/PAM hydrogel is almost the same as DS-loaded CMTKG/PVP/PAM hydrogel, which means DS shows very little or no impact on the thermal stability of the hydrogel indicating an absence of chemical interaction, but the presence of physical interaction (hydrogen bonding) between DS and the polymeric network.

The SEM images of CMTKG/PVP/PAM (A-15) hydrogel and DS-loaded CMTKG/PVP/PAM (A-5) hydrogel are displayed in Fig. 7. The

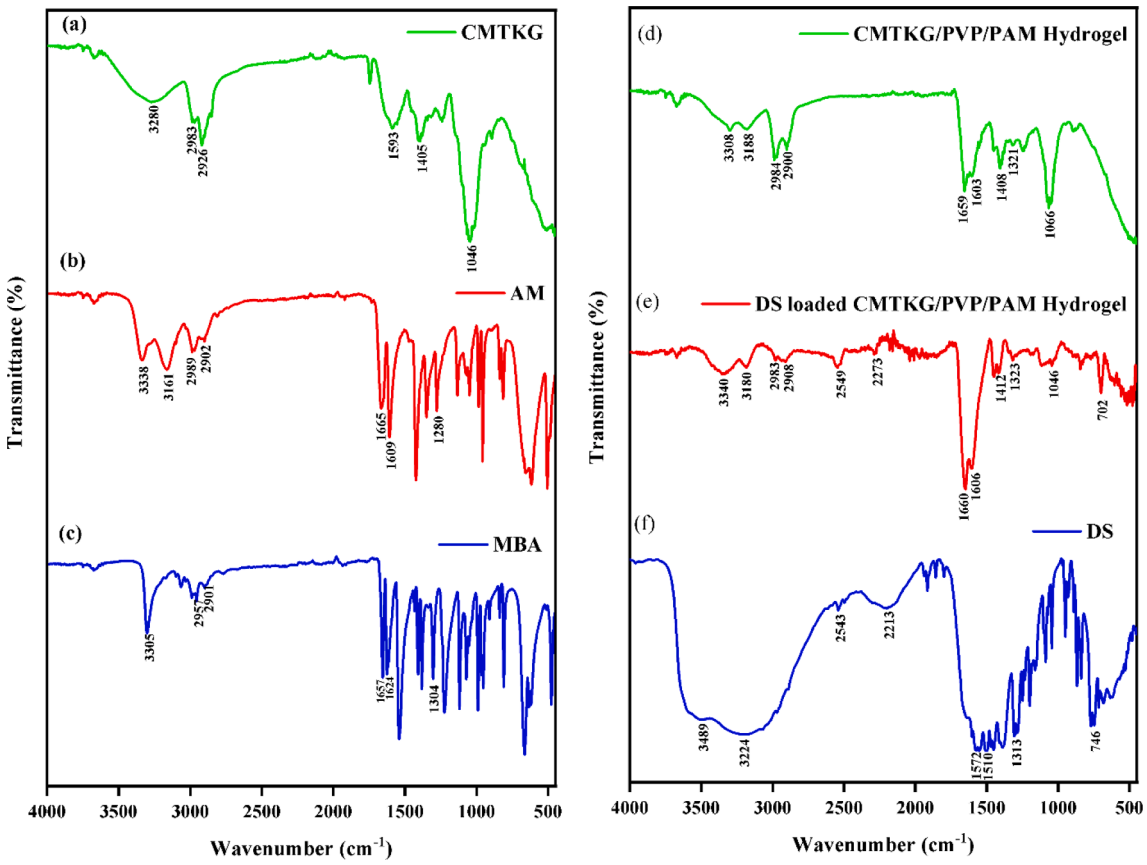


Fig. 5. ATR-FTIR spectrum of (a) CMTKG, (b) AM, (c) MBA, (d) CMTKG/PVP/PAM hydrogel, (e) DS-loaded CMTKG/PVP/PAM hydrogel, and (f) DS.

Table 2
ATR-FTIR data of (a) CMTKG, (b) AM, (c) MBA, (d) CMTKG/PVP/PAM hydrogel, (e) DS-loaded CMTKG/PVP/PAM hydrogel and (f) DS.

Functionality	Wavenumber (cm^{-1})						Ref.
	AM	MBA	CMTKG	DS	CMTKG/PVP/PAM Hydrogel	DS-loaded CMTKG/PVP/PAM Hydrogel	
C-N	1280	1304	–	1313	1321	1323	[43]
C = C and N-H bending overlap	1609	1624	–	1510	1659	1660	[43]
C = O	1665	1657	–	1572	1603	1606	[34]
OH stretching of carboxylic acid	–	–	3280	–	3300 & 3188	3340 & 3180	[36]
COO^- stretching	–	–	1405	–	1408	1412	[36]
C-H stretching	2902 & 2989	2901 & 2957	2926 & 2983	–	2900 & 2984	2908 & 2983	[42]
C-O-C	–	–	1046	–	1066	1046	[42]
C-Cl	–	–	–	746	–	702	[37]
N-H stretching	3338 & 3161	3305	–	3224 & 3489	–	–	[47]

micrographs of CMTKG/PVP/PAM hydrogel reveal uneven surface morphology, and porous structure, indicating good water retention, and drug loading capacity of the hydrogel. However, micrographs of DS-loaded CMTKG/PVP/PAM hydrogel show a smooth surface with lesser pores. Hence, the drug got incorporated into the hydrogel matrix, therefore lesser interstitial void remains in the hydrogel [30].

3.4. Drug loading and entrapment efficiency

The percent drug loading (DL%) & entrapment efficiency (DEE%) was determined for DS-loaded CMTKG/PAM hydrogel (A-14) and DS-loaded CMTKG/PVP/PAM efficient swelling hydrogel (A-5) as illustrated in Table 3. It was noticed that DS-loaded CMTKG/PVP/PAM hydrogel showed greater DL% and DEE% in comparison to DS-loaded CMTKG/PAM hydrogel. This is due to the presence of aryl ring in the PVP molecule, creating more space or enlargement taking place in the network of DS-loaded CMTKG/PVP/PAM hydrogel as well as due to the hydrophobic-hydrophobic interaction existing among the alkyl group of PVP and the aryl ring of diclofenac sodium, which enhances the loading of the sparingly water-soluble diclofenac sodium.

3.5. In vitro release study of DS

In vitro release studies were performed for DS-loaded CMTKG/PAM hydrogel (sample code A-14) and DS-loaded CMTKG/PVP/PAM hydrogel (sample code A-5) in a simulated physiological buffer solution of pH 1.2 and 7.4, at 37°C for 24 hr, presented in Fig. 8. For both DS-loaded CMTKG/PAM and DS-loaded CMTKG/PVP/PAM hydrogel, the higher

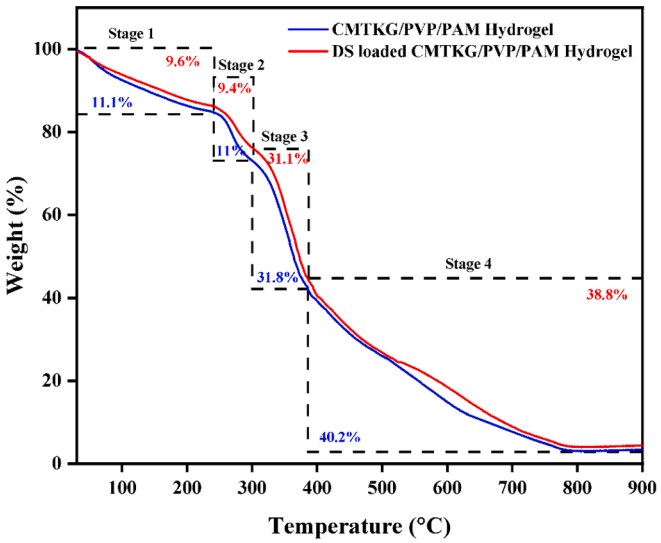


Fig. 6. TGA thermogram of CMTKG/PVP/PAM hydrogel and DS-loaded CMTKG/PVP/PAM (A-5) hydrogel.

percent of DS release was noticed at pH 7.4 because of the deprotonated COO^- anion of the CMTKG. Here, anionic-anionic repulsion among the carboxylate group causes an enlargement in network gaps. Therefore, the swelling ratio gets increases, which leads to rapid diffusion of diclofenac sodium from the polymeric matrix. However, at acidic pH, the carboxylate ion gets protonated, which results in the hydrogen bonding between PAM and CMTKG. As an outcome, shrinkage of the hydrogel took place, resulting decrease in swelling percentage as well as drug release [35]. The DS maximum release was reported to be 66.3% in DS-loaded CMTKG/PAM and 87.3% in DS-loaded CMTKG/PVP/PAM hydrogel at pH-7.4, while 22.9% and 26.3% in DS-loaded CMTKG/PAM and DS-loaded CMTKG/PVP/PAM hydrogel at pH-1.2. It was perceived that CMTKG/PVP/PAM hydrogel showed more excellent drug release compared to DS-loaded CMTKG/PAM hydrogel at pH 7.4 and 1.2. This is due to the presence of aryl ring in the PVP molecule, creating more space or enlargement taking place in the network and hydrophobic-hydrophobic interaction among the alkyl group of PVP and the aryl ring of diclofenac sodium, which causes a rise in the loading and release of the sparingly soluble diclofenac sodium. Hence, it can be concluded that the synthesized CMTKG/PVP/PAM hydrogel can be utilized as excellent material for the efficient delivery of sparingly soluble DS.

3.6. Release kinetic modeling of diclofenac sodium

The higher percent of drug release was noticed in the case of DS-loaded CMTKG/PVP/PAM hydrogel (A-5). Therefore, kinetic modeling was performed for A-5 hydrogel, using Korsmeyer-Peppas, First-order, Higuchi, Zero-order, and Hixson-Crowell models, as presented in Table 4. Based on the R^2 value for all the models, the best suitable model to explain the drug release mechanism was selected. The R^2 value for the Korsmeyer-Peppas model was the highest of the other models. The release exponent "n" suggested the type of diffusion. The R^2 value based on the Korsmeyer-Peppas model at pH 1.2 is found to be 0.9954, and the diffusion exponent value (n) value is 0.4848. In pH 7.4, the R^2 was 0.9931, and the n value was 0.3976. The kinetic graphs for this model are shown in Fig. 9. However, based on the n value, the diffusion mechanism was observed as Fickian at both pH values, which implies that drug release from the polymeric network is solely diffusion dependent.

3.7. Conclusion

This research work reports the successful synthesis of smart diclofenac sodium-loaded CMTKG-based hydrogel that can be used for the controlled release of sparingly soluble diclofenac sodium. Furthermore, the impact of the concentration of initiator, biopolymer, and crosslinker on the swelling nature of synthesized hydrogel was also studied. The swelling and drug release was examined in alkaline and acidic pH and found to be higher in alkaline condition, attributed to strong electrostatic repulsion between COO^- anion of the CMTKG. Sol-gel analysis was conducted by varying amounts of CMTKG, MBA, or KPS and shows an increase in gel fraction with the increase in the amount of reactant. The effect of the addition of PVP on drug loading and release was studied by comparing formulation (A-5), which contains PVP to formulation (A-14,) which lacks PVP. The result indicates that in the presence of amphiphilic polymer (PVP), a hydrophobic character imparts in the polymeric matrix, leading to increased drug loading and release capacity. The kinetics of drug release were effectively correlated with the

Table 3
DL% and DEE% of DS-loaded CMTKG/PAM and DS-loaded CMTKG/PVP/PAM (A-5) hydrogel.

Formulation	Hydrogel	Drug Loading (DL%)	Drug Entrapment Efficiency (DEE%)
A-14	DS-loaded CMTKG/PAM	17.25	60.35
A-5	DS-loaded CMTKG/PVP/PAM	20.34	65.54

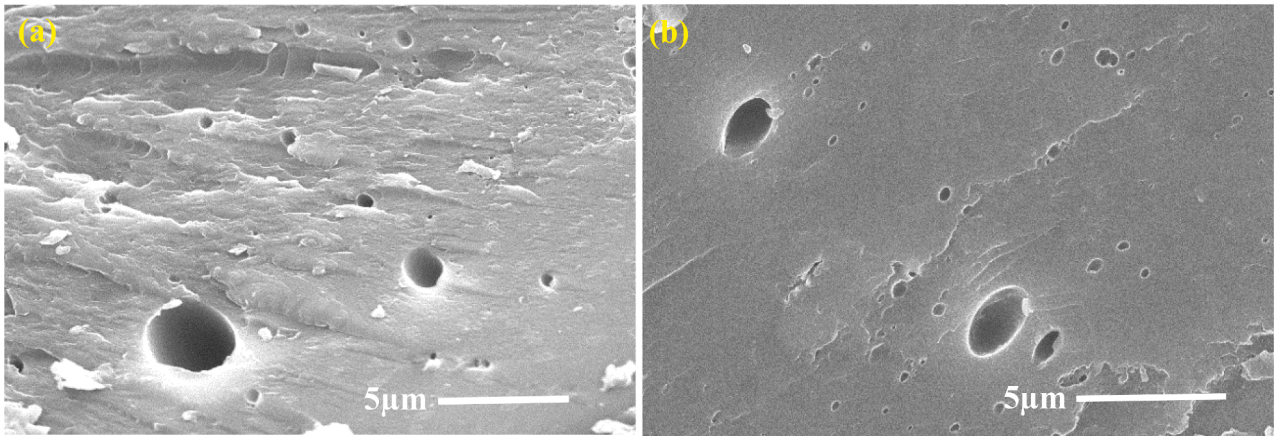


Fig. 7. SEM image of (a) CMTKG/PVP/PAM hydrogel and (b) DS-loaded CMTKG/PVP/PAM (A-5) hydrogel.

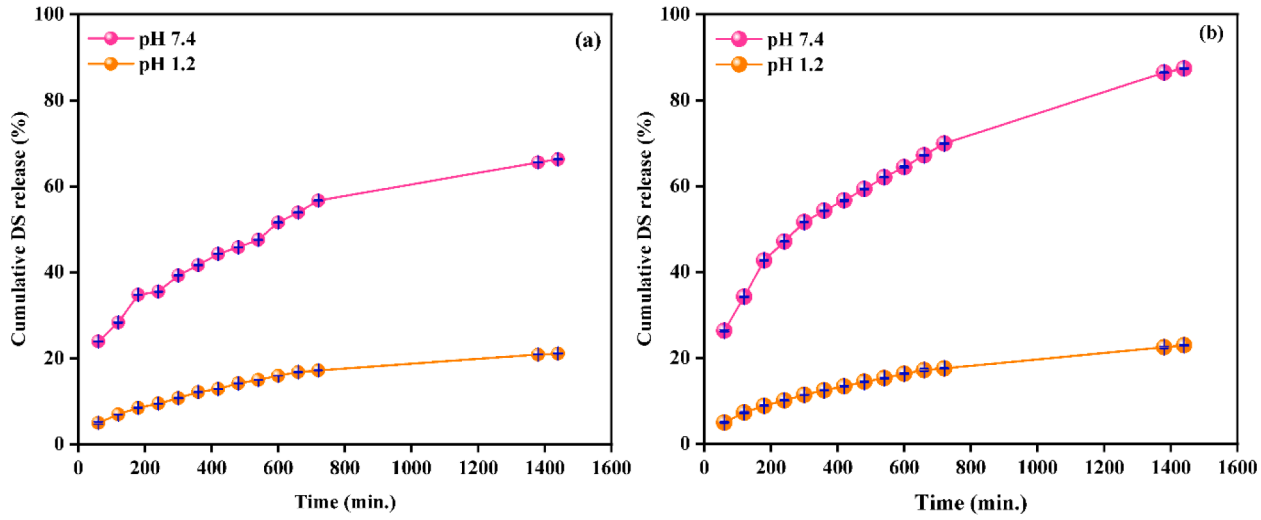


Fig. 8. Cumulative drug release curve for (a) DS-loaded CMTKG/PAM hydrogel and (b) DS-loaded CMTKG/PVP/PAM (A-5) hydrogel in pH 7.4 and 1.2.

Table 4
Various models followed for the release kinetic studies of DS from CMTKG/PVP/PAM hydrogel (A-5).

Model	Mechanism	Equation	pH 7.4		pH 1.2		Ref.
			n	R^2	n	R^2	
Zero Order	Rate is independent of the amount of the drug	$M_t = M_\infty + k_0 t$	–	0.8989	–	0.9111	[48]
Higuchi	Fickian diffusion followed	$F = M_t/M_\infty = k_H t^{½}$ k_H = kinetic constant	–	0.9862	–	0.9907	[37]
First Order	The drug release rate is directly proportional to the quantity of the drug.	$\text{Log } M_t = \text{Log } M_\infty + \frac{kt}{2.303}$ k = rate constant	–	0.7563	–	0.7476	[49]
Hixson-Crowell	Erosion	$(M_t)^{1/3} - (M_\infty)^{1/3} = k_{HC} t$ k_{HC} = Hixson Crowell constant	–	0.8096	–	0.8109	[49]
Korsmeyer–Peppas	1. Fickian Diffusion (n < 0.5)2. Non-Fickian Diffusion (0.89 ≥ n ≥ 0.5) 3. Case II transport (n > 0.89)	$F = M_t/M_\infty = kt^n$ k = kinetic constant n = diffusion exponent	0.3976	0.9931	0.4848	0.9954	[50]

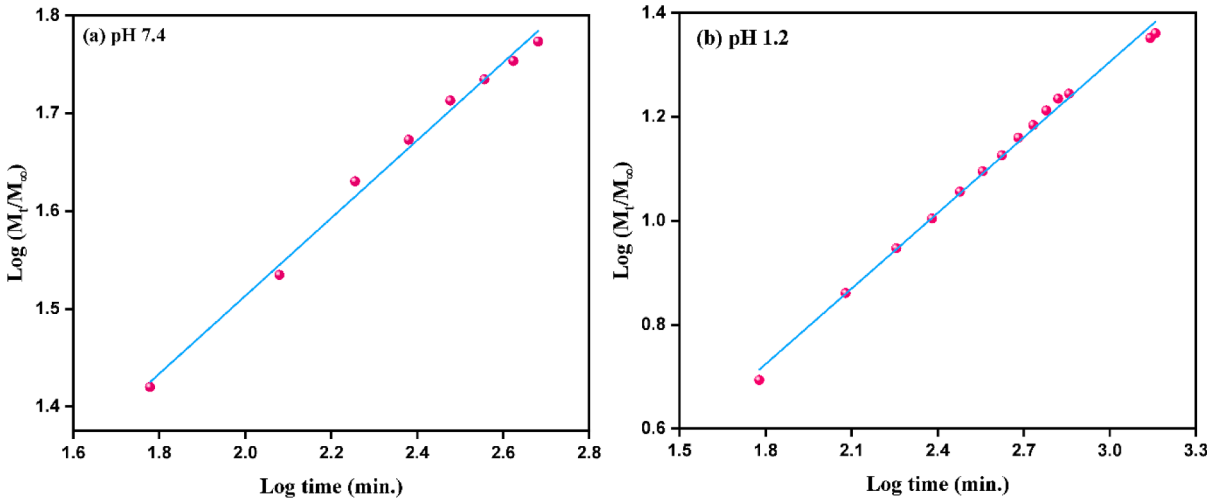


Fig. 9. Release kinetic of DS from CMTKG/PVP/PAM (A-5) hydrogel based on the Korsmeyer–Peppas model in **(a)** pH-7.4, **(b)** 1.2.

Korsmeyer-Peppas model, suggesting drug release from the hydrogel is purely diffusion dependent as it follows the Fickian diffusion mechanism. The above findings permit the utilization of DS-loaded CMTKG/PVP/PAM hydrogel as a feasible material for the pH-dependent release of sparingly soluble diclofenac sodium.

CRediT authorship contribution statement

Priyanka Meena: Conceptualization, Methodology, Visualization, Resources, Writing – original draft, Formal analysis, Investigation, Data curation. **Poonam Singh:** Writing – review & editing, Supervision. **Sudhir G. Warkar:** Conceptualization, Validation, Project administration, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank the Delhi Technological University, New Delhi, India, for its infrastructure and the Council of Scientific & Industrial Research (CSIR) for providing Senior Research Fellowship (SRF).

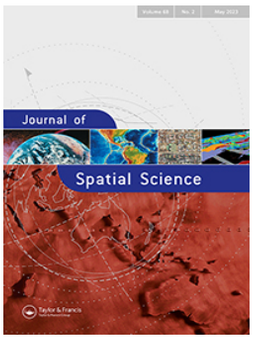
Data and code availability

Not applicable.

References

[1] X. Bi, H. Zhang, L. Dou, Layered double hydroxide-based nanocarriers for drug delivery, Pharmaceutics 6 (2014) 298–332, https://doi.org/10.3390/pharmaceutics6020298.
[2] B.K. Nanjwade, G.M. Hiremath, Biopolymers for drug delivery, Springer International Publishing, New York, 2011, pp. 655–670.
[3] A.V.V.V. Ravi Kiran, G. Kusuma Kumari, P.T. Krishnamurthy, Carbon nanotubes in drug delivery: Focus on anticancer therapies, J. Drug Deliv. Sci. Technol. 59 (2020), 101892, https://doi.org/10.1016/j.jddst.2020.101892.
[4] L. Liu, Q. Ma, J. Cao, et al., Recent progress of graphene oxide-based multifunctional nanomaterials for cancer treatment, Cancer Nanotechnol. 12 (2021) 1–31. https://doi.org /10.1186/s12645-021-00087-7.
[5] S. Senapati, A.K. Mahanta, S. Kumar, P. Maiti, Controlled drug delivery vehicles for cancer treatment and their performance, Signal Transduction and Targeted Therapy 3 (2018) 1–19, https://doi.org/10.1038/s41392-017-0004-3.
[6] R. Bhattacharyya, P. Chowdhury, Hydrogels of Acryloyl guar gum–g–(acrylic acid–co–3sulfopropylacrylate) for high–performance adsorption and release of gentamicin sulphate, J. Polym. Res. 28 (2021) 286, https://doi.org/10.1007/s10965-021-02633-8.
[7] P. Chowdhury, A. Mukherjee, B. Singha, Synthesis of cross-linked graft copolymer hydrogel from maleic acid, poly(vinyl alcohol) and methylene bisacrylamide, and studies on its efficiency in uptaking zinc ion from aqueous solution, J. Indian Chem. Soc. 88 (2011) 1517-1524. https: //doi.org/ 10.5281/zenodo. 5790609.
[8] I. Gholamali, M. Yadollahi, Bio-nanocomposite polymer hydrogels containing nanoparticles for drug delivery: a review, Regen Eng. Transl. Med. 7 (2021) 129–146, https://doi.org/10.1007/s40883-021-00207-0.
[9] K. Hemmati, R. Sahraei, M. Ghaemy, Synthesis and characterization of a novel magnetic molecularly imprinted polymer with incorporated graphene oxide for drug delivery, Polymer 101 (2016) 257–268, https://doi.org/10.1016/j.polymer.2016.08.074.
[10] K. Hemmati, A. Masoumi, M. Ghaemy, Synthesis and characterization of pH-responsive nanohydrogels as biocompatible drug carriers based on chemically modified tragacanth gum polysaccharide, RSC Adv. 104 (2015) 85310–85318, https://doi.org/10.1039/c5ra14356j.
[11] S.A. Munim, Z.A. Raza, Poly(lactic acid) based hydrogels: formation, characteristics and biomedical applications, J. Porous Mater. 26 (2019) 881–901, https://doi.org/10.1007/s10934-018-0687-z.
[12] K. Hemmati, M. Ghaemy, Synthesis of new thermo/pH sensitive drug delivery systems based on tragacanth gum polysaccharide, Int. J. Biol. Macromol. 87 (2016) 415–425, https://doi.org/10.1016/j.ijbiomac.2016.03.005.
[13] A. Martínez-Ruvalcaba, J.C. Sánchez-Díaz1, F. Becerra, et al., Swelling characterization and drug delivery kinetics of polyacrylamide-co-itaconic acid/chitosan hydrogels, eXPRESS Polym. Lett. 3 (2009) 25–32. doi:10.3144/expresspolymlett. 2009.5.
[14] K. Thirupathi, T.T.V. Phan, M. Santhamoorthy, V. Ramkumar, S.-C. Kim, pH and thermoresponsive PNIPAm-co-polyacrylamide hydrogel for dual stimuli-responsive controlled drug delivery, Polymers 15 (2022) 167.
[15] H. Musa, S. Yahaya, Polyacrylamide hydrogels for application in oral drug delivery, Niger J. Sci. Res. 20 (2022) 390–396.
[16] B. Biman, Mandal, Sonia Kapoor, Subhas C. Kundu, Silk fibroin/polyacrylamide semi-interpenetrating network hydrogels for controlled drug release, Biomaterials 30 (2009) 2826–2836, https://doi.org/10.1016/j.biomaterials.2009.01.040.
[17] S. Nangia, S. Warkar, D. Katyal, A review on environmental applications of chitosan biopolymeric hydrogel based composites, J. Macromol. Sci. Part A Pure Appl. Chem. 55 (2018) 747–763, https://doi.org/10.1080/10601325.2018.1526041.
[18] M. Hassanzadeh, M. Ghaemy, S. Ahmadi, Extending time profile of morphine-induced analgesia using a chitosan-based molecular imprinted polymer nanogel, Macromol. Biosci. 16 (2016) 1515–1523, https://doi.org/10.1002/mabi.201600177.
[19] K. Hemmati, A. Masoumi, M. Ghaemy, pH responsive tragacanth gum and poly (methyl methacrylate-co-maleic anhydride)-g-poly(caprolactone) conetwork microgel for in vitro quercetin release, Polymer 59 (2015) 49–56. doi:10.1016/ j. polymer. 2014.12.050.
[20] P. Goyal, V. Kumar, P. Sharma, Carboxymethylation of Tamarind kernel powder, Carbohydr. Polym. 69 (2007) 251–255, https://doi.org/10.1016/j.carbpol.2006.10.001.
[21] S. Pal, G. Sen, S. Mishra, R.K. Dey, U. Jha, Carboxymethyl tamarind: Synthesis, characterization and its application as novel drug-delivery agent, J. Appl. Polym. Sci. 110 (2008) 392–400.
[22] Khushbu, S.G. Warkar, A. Kumar, Synthesis and assessment of carboxymethyl tamarind kernel gum based novel superabsorbent hydrogels for agricultural applications, Polymer 182 (2019) 121823.
[23] S. Jana, A. Banerjee, K.K. Sen, S. Maiti, Gelatin-carboxymethyl tamarind gum biocomposites: In vitro characterization & anti-inflammatory pharmacodynamics, Mater. Sci. Eng. C 69 (2016) 478–485, https://doi.org/10.1016/j.msec.2016.07.008.
[24] K. Mali, S. Dhawale, R. Dias, V. Ghorpade, Delivery of drugs using tamarind gum and modified tamarind gum: A review, Bull Fac Pharmacy Cairo Univ 57 (2019) 1–24. doi:10.21608/BFPC.2019.47260.
[25] E. Larrañeta, S. Stewart, M. Ervine, et al., Hydrogels for hydrophobic drug delivery, classification, synthesis and applications, J. Funct. Biomater. 9 (2018) 13.
[26] Khushbu, S.G. Warkar, Potential applications and various aspects of polyfunctional macromolecule- carboxymethyl tamarind kernel gum, Eur. Polym. J. 140 (2020) 110042.
[27] P. Ghasemiyeh, S. Mohammadi-Samani, Hydrogels as Drug Delivery Systems; Pros and Cons., Trends Pharm Sci 5 (2019) 7–24. 10.30476/TIPS.2019. 81604.1002.
[28] D.-K. Lim, M.-H. Cui, J.-M. Nam, Highly stable, amphiphilic DNA-encoded nanoparticle conjugates for DNA encoding/decoding applications, J. Mater. Chem. 21 (2011) 9467.
[29] M. Kurakula, G.S.N. Koteswara Rao, Pharmaceutical assessment of polyvinylpyrrolidone (PVP): As excipient from conventional to controlled delivery systems with a spotlight on COVID-19 inhibition, J. Drug Delivery Sci. Technol. 60 (2020), 102046, https://doi.org/10.1016/j.jddst.2020.102046.
[30] SE Saraydin, Interpenetrating polymeric network hydrogels for potential gastrointestinal, Polym Int 56 (2020) 1371–1377, https://doi.org/10.1002/pi.2271.
[31] H. Shoukat, F. Pervaiz, S. Noreen, M. Nawaz, R. Qaiser, M. Anwar, Fabrication and evaluation studies of novel polyvinylpyrrolidone and 2-acrylamido-2-methylpropane sulphonic acid-based crosslinked matrices for controlled release of acyclovir, Polym. Bull. 77 (2020) 1869–1891.
[32] A.K. Nayak, S. Khatua, M.S. Hasnain, K.K. Sen, Development of diclofenac sodium-loaded alginate-PVP K 30 microbeads using central composite design, DARU, J. Pharm. Sci. 19 (2011) 356–366.
[33] M. Suhail, A. Khan, J.M. Rosenholm, M.U. Minhas, P.-C. Wu, Fabrication and characterization of diclofenac sodium loaded hydrogels of sodium alginate as sustained release carrier, Gels 7 (2021) 10.
[34] H. Hosseinzadeh, Controlled release of diclofenac sodium from pH-responsive carrageenan-g-poly(acrylic acid) superabsorbent hydrogel, J. Chem. Sci. 122 (2010) 651–659, https://doi.org/10.1007/s12039-010-0100-1.
[35] Q.-B. Wei, F. Fu, Y.-Q. Zhang, Q. Wang, Y.-X. Ren, PH-responsive CMC/PAM/PVP semi-IPN hydrogels for theophylline drug release, J. Polym. Res. 21 (2014), https://doi.org/10.1007/s10965-014-0453-0.
[36] D. Chauhan, A. Kumar, S.G. Warkar, Synthesis, characterization and metal ions sensing applications of meta- benziporphodimethene-embedded polyacrylamide/carboxymethyl guargum polymeric hydrogels in water ENVIRONMENTAL, TECHNOLOGY (2020), https://doi.org/10.1080/09593330.2020.1812730.
[37] Khushbu, S.G. Warkar, N. Thombare, Controlled release and release kinetics studies of boron through the functional formulation of carboxymethyl tamarind kernel gum-based superabsorbent hydrogel, Polym. Bull. 79 (2022) 2287–2303.
[38] L.S. Tan, H.L. Tan, K. Deekonda, Y.Y. Wong, S. Muniyandy, K. Hashim, J. Pushpamalar, Fabrication of radiation cross-linked diclofenac sodium loaded carboxymethyl sago pulp/chitosan hydrogel for enteric and sustained drug delivery, Carbohydr. Polym. Technol. Appl. 2 (2021) 100084.
[39] S. Nawaz, S. Khan, U. Farooq, M.S. Haider, N.M. Ranjha, A. Rasul, A. Nawaz, N. Arshad, R. Hameed, Biocompatible hydrogels for the controlled delivery of anti-hypertensive agent: Development, characterization and in vitro evaluation, Des. Monomers Polym. 21 (2018) 18–32.
[40] B. Niu, J. Jia, H. Wang, S. Chen, W. Cao, J. Yan, X. Gong, X. Lian, W. Li, Y.-Y. Fan, In vitro and in vivo release of diclofenac sodium-loaded sodium alginate/carboxymethyl chitosan-ZnO hydrogel beads, Int. J. Biol. Macromol. 141 (2019) 1191–1198.
[41] S. Jana, R. Sharma, S. Maiti, K.K. Sen, Interpenetrating hydrogels of O-carboxymethyl Tamarind gum and alginate for monitoring delivery of acyclovir, Int. J. Biol. Macromol. 92 (2019) 1034–1039, https://doi.org/10.1016/j.ijbiomac.2016.08.017.
[42] M.L. Varahala Setti, J.V. Ratna, Preparation and evaluation of controlled release tablets of carvedilol, Asian J. Pharm. 3 (2009) 252–256, https://doi.org/10.4103/0973-8398.56307.
[43] R. Malik, S.G. Warkar, R. Saxena, Carboxy-methyl tamarind kernel gum based bio-hydrogel for sustainable agronomy, Mater. Today Commun. 35 (2023), 105473, https://doi.org/10.1016/j.mtcomm.

[44] K.K. Mali, S.C. Dhawale, R.J. Dias, Synthesis and characterization of hydrogel films of carboxymethyl tamarind gum using citric acid, Int. J. Biol. Macromol. 105 (2017) 463–470, https://doi.org/10.1016/j.ijbiomac.2017.07.058.
[45] E.M. Elzayat, A.A. Abdel-Rahman, S.M. Ahmed, et al., Studying the impact of formulation and processing parameters on the release characteristics from hydroxypropyl methylcellulose matrix tablets of diclofenac, Acta Pol. Pharm. - Drug Res. 73 (2016) 439–452.
[46] K.S.V.K. Rao, M.C.S. Subha, M. Sairam, et al., Synthesis, characterization and controlled release characteristics of PEGylated hydrogels for diclofenac sodium, Des. Monomers Polym. 9 (2006) 261–273, https://doi.org/10.1163/15685550677735099.2.
[47] R. Abdelkader, B. Mohammed, Green synthesis of cationic polyacrylamide composite catalyzed by an ecologically catalyst clay called maghnite-H^+ (Algerian MMT) under microwave irradiation, Bull. Chem. React. Eng. Catal. 11 (2016) 170–175, https://doi.org/10.9767/bcrec.11.2.543.170-175.
[48] N. Goonoo, A. Bhaw-Luximon, R. Ujoodha, A. Jhugroo, G.K. Hulse, D. Jhurry, Naltrexone: A review of existing sustained drug delivery systems and emerging nano-based systems, J. Control. Release 183 (2014) 154–166.
[49] D. Das, R. Das, J. Mandal, A. Ghosh, S. Pal, Dextrin crosslinked with poly(lactic acid): A novel hydrogel for controlled drug release application, J. Appl. Polym. Sci. 131 (2014), https://doi.org/10.1002/app.40039.
[50] Tushar, Y. Saraswat,P. Meena, S.G. Warkar, Synthesis and characterization of novel xanthan gum-based pH-sensitive hydrogel for metformin hydrochloride release, Colloid Polym. Sci. (2023), doi:10.1007/s00396-023-05135-9.



Journal of Spatial Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjss20

Development of flood susceptibility map using a GIS-based AHP approach: a novel case study on Idukki district, India

Zohaib Ahmed Khan & Bharat Jhamnani

To cite this article: Zohaib Ahmed Khan & Bharat Jhamnani (2023): Development of flood susceptibility map using a GIS-based AHP approach: a novel case study on Idukki district, India, Journal of Spatial Science, DOI: 10.1080/14498596.2023.2236051

To link to this article: https://doi.org/10.1080/14498596.2023.2236051

Published online: 02 Aug 2023.

Submit your article to this journal

Article views: 24

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tjss20

JOURNAL OF SPATIAL SCIENCE
https://doi.org/10.1080/14498596.2023.2236051

Taylor & Francis
Taylor & Francis Group

Check for updates

Development of flood susceptibility map using a GIS-based AHP approach: a novel case study on Idukki district, India

Zohaib Ahmed Khan and Bharat Jhamnani

Department of Civil Engineering, Delhi Technological University, Delhi, India

ABSTRACT

Flooding is a regular danger in the Idukki district of Kerala causing serious damage to life and property. An attempt has been undertaken in this research to identify crucial zones in the Idukki district that are more susceptible to flooding. The AHP technique was used to give relative weights to 12 flood contributing parameters. The resultant flood susceptibility map shows that 1346.4 km^2 of the total land in the Idukki district has a high to very high likelihood of flooding. Validation of the resultant map was achieved using the AUC-ROC approach, which showed good accuracy of 0.91.

ARTICLE HISTORY
Received 14 March 2023
Accepted 9 July 2023

KEYWORDS
Floods; analytical hierarchy process; susceptibility; mapping; remote sensing

1. Introduction

Flooding is a regular natural hazard that causes significant damage to people and property (Gabriels *et al*. 2022). Floods benefit the ecosystem in a number of ways at lower magnitudes, including by supplying water and nutrients to riparian corridors, eliminating pollutants from flood plain zones, recharging the groundwater table and enhancing soil fertility (Lyubimova *et al*. 2016, Zhang *et al*. 2017, Kraus *et al*. 2019, Abijith *et al*. 2020, Mahato *et al*. 2021). The lives, agricultural operations, residential areas, properties, roadways and natural ecosystems will all suffer enormous damage if the flood tolerance limit is exceeded (Pradhan and Youssef 2011, George *et al*. 2022). According to Kron (2005), this occurs due to heavy rain and melting of snow, which causes the river to overflow and stay momentarily stagnant, causing damage to the lands along the river. Flooding can be caused due to the overflowing of a river as well as other factors like abrupt cloudbursts in semi-arid locations, resulting in flash flood (Tehrany *et al*. 2013). Flooding is a very complicated occurrence that has always drawn researchers from all around the world to better understand and study different mechanisms for better management and avoidance of flood-related hazards. Several natural and human factors are to blame, all of which have the potential to cause a catastrophic flood disaster (Sholihah *et al*. 2020). There is an increase in frequencies of flood episodes and the causes are often associated with unsustainable industrial and farming practices such as cutting of trees, excessive utilisation of resources, improper planning of urbanised colonies and concretisation (Bradshaw *et al*. 2007).

CONTACT Zohaib Ahmed Khan zohaib098765@gmail.com
© 2023 Mapping Sciences Institute, Australia and Geospatial Council of Australia

Climate change has also been a significant impact in recent catastrophic flood episodes (Sampson *et al*. 2015). Climate-induced changes, according to (Charlton *et al*. 2006), can greatly affect an area's flood risk by changing land use patterns thereby creating an impervious surface, which increases the velocity of water. As the hydrological cycle becomes more intense due to climate change, flood-related hazards are anticipated to grow, posing a risk to people and property all over the world. Aside from climate, other elements influence flood occurrence, including a region's altitudinal features, vicinity to the main channel, slope, topographic curvature and geological variations (Lim and Lee 2017). Floods can also occur when the excess rainfall due to heavy downpour does not get routed in a particular time frame, which ultimately results in waterlogging of the region. Studies also indicate that increase in the flooding events over the years is due to the rapid increase in artificial surfaces such as roads and buildings in urban areas. Such artificial surfaces are highly impervious and do not allow water to infiltrate into the ground, which leads to floods and heavy waterlogging during events of extreme rainfall (Bronstert 2003).

India experiences a monsoon season spanning from June to September, wherein the country receives a substantial quantum of precipitation. Monsoons play a crucial role in facilitating agricultural practices; however, they can result in an overabundance of precipitation and consequent inundation (Saha *et al*. 2021). A number of regions in India are susceptible to flooding owing to various factors including but not limited to intense precipitation, insufficient drainage infrastructure, deforestation, urbanisation and suboptimal land utilisation practices. The management of floods in India is faced with additional challenges due to the effects of climate change and the swift process of urbanisation (Neumann *et al*. 2015). The alteration of precipitation distribution, increased magnitude of precipitation occurrences and intrusion upon indigenous water reservoirs intensify the susceptibility to flooding. The implementation of sustainable water management practices, resilient infrastructure, land-use regulations and long-term planning are essential. Several regions in India are highly susceptible to flooding, including Assam, Bihar, Uttar Pradesh, West Bengal, Odisha, Maharashtra, Kerala and Karnataka (Parthasarathy *et al*. 2021). Annually, floods in India have a significant impact on a vast number of individuals, reaching into the millions. Natural hazards have the potential to cause fatalities, displace communities, inflict harm on residential structures, devastate crops, impede transportation and communication infrastructure and propagate waterborne illnesses. The groups that are most susceptible to adverse impacts are those that are marginalised and reside in regions with low elevation or in close proximity to river banks (Abijith and Saravanan 2021). The occurrence of floods in India has noteworthy economic implications. The adverse consequences of this phenomenon extend to the structural integrity of various forms of infrastructure, including roads, bridges, railways and buildings, resulting in significant economic repercussions. The inundation of fields due to flooding can have a detrimental impact on the agriculture sector, leading to the destruction of crops and livestock, which in turn can have adverse effects on food production and the livelihoods of those involved in the industry. Industries situated in regions susceptible to flooding also encounter interruptions to their operations (Pramanick *et al*. 2022). The Indian government, at both the central and state levels, has been engaged in the development of flood management strategies and disaster response mechanisms. The endeavours encompass the establishment of flood control infrastructure, such as embankments, dams and reservoirs, along with the implementation of early-warning systems, rescue operations and relief measures for impacted

communities. The National Disaster Management Authority (NDMA) and State Disaster Management Authorities (SDMAs) are pivotal entities in the realm of disaster preparedness and response (Parthasarathy *et al*. 2021).

In recent years, India has had some of the most unusual severe precipitation occurrences, which have resulted in flooding and the loss of lives. For example, the Kosi River, commonly known as the 'Sorrow of Bihar', is renowned for its catastrophic floods in Bihar. The river in question has its source in Nepal and subsequently flows into Bihar, where it has been known to alter its course over time, resulting in devastating inundations. Due to the substantial amount of sediment transported by the river, it is susceptible to obstructions and embankment breaches, resulting in extensive flooding. The heavy precipitation in Uttarakhand in 2013 led to massive floods, killing over 6000 people and causing an economic loss of over 3.8 billion dollars ('Rapidly Assessing Flood Damage in Uttarakhand, India' 2013). The floods were predominantly triggered by extensive and uninterrupted precipitation throughout the monsoon period. The susceptibility of the region to landslides and flash floods was attributed to the combination of steep slopes, fragile geology and intense rainfall. According to Houze *et al*. (2017) the main cause of the disastrous flooding in Uttarakhand, India in June 2013 was the condition of deep convection which moistened the mountainsides, thereby making the region more vulnerable to floods. Flooding in Chennai was triggered by a major rain storm in 2015, which caused $3 billion in damage (Van Oldenborgh *et al*. 2016). The state of Assam experiences recurring floods on a yearly basis, resulting in substantial damage and displacement. Floods are typified by the inundation of extensive regions, the erosion of riverbanks, the breaching of embankments and the overflowing of rivers. The Brahmaputra River presents a considerable flood hazard owing to its substantial discharge and ever-changing characteristics. In India and other parts of the world, extreme rainfall occurrences and flooding have become among the most expensive natural catastrophes (The Human Cost of Natural Disasters 2015: A Global Perspective 2015). Flooding-related human losses are expected to increase by 70–80% if the global mean temperature rises over 1.5°C above pre-industrial levels (Dottori *et al*. 2018).

Catastrophic rainfall and extensive flooding in Kerala recently demonstrated the magnitude of India's extreme rainfall and large-scale floods. In August 2018, Kerala saw prolonged and excessive rains that impacted many elements of human life, including socioeconomic circumstances, transportation, infrastructure, agriculture and livelihood. The Kerala floods of 2018, which are probably the worst floods in a century, have already drawn the attention of the media, the scientific community and policymakers (Hunt and Menon 2020). According to preliminary estimates, the Kerala floods killed over 440 people and inflicted $3 billion of economic damage (Young *et al*. 2022). According to the India Meteorological Department (IMD), Kerala recorded the highest daily rainfall in six decades in 2021. This resulted in excess rainfall which flooded most of the agricultural areas, due to which crop production was adversely affected, which further impacted the livelihoods of farmers.

In recent decades, remote sensing and geographic information systems (GIS) have grown as instruments for management of natural hazards and geospatial mapping by employing different multi-criteria decision techniques (Das 2017, 2018). For instance, Chen *et al*. (2011) delineated flood susceptibility zones in the cities of Taiwan with the help of GIS coupled with the AHP approach. Kazakis *et al*. (2015) employed the AHP approach for delineating flood susceptible maps in Rhodope-Evros, which is a region in

Greece. The developed flood susceptible maps were validated with historical flood occurrence. Hsu *et al*. (2017) utilised the AHP and GIS approach to map flood-prone areas in the Fangshan district, China, employing nine flood conditioning factors, where six factors namely rainfall intensity, rainfall frequency, slope, elevation, river network density and impervious surface were considered in relation to hazard and three factors namely population, road network density and fixed assets were considered for vulnerability. The resultant map was successfully validated with the locations of historical flood events. Tran *et al*. (2017) developed flood susceptible zones in coastal lowland regions of central Vietnam in which they showed that factors such as elevation and distance to river play a significant part in the development of the final flood susceptibility map. Seejata *et al*. 2018 employed the AHP model for mapping of flooded regions of Sukhothai province in Thailand utilising six parameters, namely elevation, slope, rainfall intensity, land use patterns, soil permeability and channel density. Hamlat *et al*. (2021) employed the AHP method and remote sensing data along with the integration of GIS, and mapped the flood-hazard-prone regions in M'zi wadi basin in Algeria using seven parameters, namely elevation, rainfall intensity, flow accumulation, slope of the basin, drainage density, land use/land cover and type of soil present in the basin. The developed flood hazard map was successfully validated with help of pre-flood and post-flood satellite images of the basin.

Koks *et al*. (2015) utilised combined hazard, exposure and social vulnerability to enhance the detailed assessment of flood risk management strategies. Hadipour *et al*. (2020) employed GIS-based spatial multi-criteria decision analysis and developed a flood risk index due to coastal flooding and mapped it by integrating flood hazard, social vulnerability and exposure as risk components. Paranunzio *et al*. (2022) emphasised the importance of changes in rise in sea level, urban sprawl and land cover changes in the evaluation of coastal flood risk in the county of Dublin. Ghosh and Kar (2018) stressed the importance of including socio-economic parameters along with physical aspects to assess flood hazard and vulnerability through AHP-based approaches. Chakraborty and Mukhopadhyay (2019) signified the efficiency of coupled AHP and GIS approaches as a reliable tool for spatial flood risk distribution in regions where limited data are available. Over the last two decades, various methods for studying flood-prone areas have been developed, including the AHP approach (Dou *et al*. 2017), genetic algorithms and fuzzy logic (Nourani *et al*. 2014, Hong *et al*. 2018), decision tree models, multivariate statistics (Tehrany *et al*. 2013) and a hydrological forecasting system (Abbaszadeh 2016, Danandeh Mehr and Kahya 2017).

The flood susceptibility mapping makes extensive use of a variety of machine learning methods, including random forest, CART, support vector machines, XGBoost, gradient boosting machines and artificial neural networks (ANNs). ANNs have been used extensively in the past 10 years to map the susceptibility of natural hazards (Giovannettone *et al*. 2020, Jacinth Jennifer and Saravanan 2022), but they have shortcomings like under- and over-fitting, slow learning, the curse of dimensionality and slow convergence to a local optimum; in addition, their processing of complex hydrological phenomena has been insufficient (Krishnaraj and Honnasiddaiah 2022, Saravanan and Abijith 2022). The AHP model and geospatial approaches are the simplest way to predict flood susceptible zones by carefully examining multiple relevant components (Gupta and Dixit 2022). Inaccuracy in assigning weights to the different parameters can lead to an inconsistent spatial distribution of flood risk potentials. Over the years, the AHP in connection with GIS has gained more popularity for mapping floods (Pal and Ziaul 2017, Aslam *et al*. 2022).

Also, the outcomes of GIS-based statistical calculations showed more accuracy than using spatial data alone (Farhan Ul Moazzam *et al*. 2018, Swain *et al*. 2020). Several studies (Vimal *et al*. 2018, Singh 2019, Hunt and Menon 2020) have attempted to investigate and identify the flood-prone zones. The AHP technique is an approach for making decisions that aids in selecting different factors and criteria according to their relative relevance and may be used to the creation of comprehensive flood susceptibility models (Chakraborty and Mukhopadhyay 2019). Various studies (Dou *et al*. 2017, Pal and Ziaul 2017, Ghosh and Kar 2018, Vimal *et al*. 2018, Aslam *et al*. 2022) used only limited number of parameters with comparison to rainfall data. The utilisation of a restricted set of parameters in the mapping of flood-prone areas may heighten the likelihood of producing less precise flood susceptibility maps.

In the current study, we have used 12 parameters related to elevation and other conditions connected to terrain for accurate flood susceptibility mapping. Specifically, in the Idukki region of Kerala, where floods have been a constant phenomenon in the recent past, no study has attempted flood susceptibility mapping using 12 parameters and validated the maps using pre- and post-flood satellite images. One of the most important aspects of flood disaster research is flood susceptibility mapping (Tehrany *et al*. 2017, 2019, Bui *et al*. 2019, Das 2019, Band *et al*. 2020, Rafiei-Sardooi *et al*. 2021). As Idukki has been more prone to floods in recent years, a proper flood analysis has to be carried out to find the flood susceptibility region. Therefore, it is of utmost importance to develop flood susceptible maps in the Idukki district of Kerala. As a result, the objective of this article is to utilise remote sensing and geospatial analysis to examine the various influencing elements that are responsible for the frequent occurrence of floods in Kerala's Idukki District, as well as to demarcate flood-prone regions by constructing a flood susceptibility map.

2. Study area

The study region, Idukki District, is part of the Western Ghats and is characterised by rough hills and deep valleys. It is located in the southwest part of Kerala. It covers an area of 4358 km^2 and is located between 9°20′N and 10°20′N latitudes and 76°25′E and 77°25′E longitudes (Figure 1). The average annual rainfall is between 250 and 425 cm and the temperature varies between 21° and 29°C. Geology of the research area shows that recent sediments overlie Precambrian rocks, with no significant Phanerozoic representation other than a huge unconformity. Biotite gneiss, granite gneiss, charnockite, granite, hornblende biotite gneiss, calc granulite and quartzite are the Precambrian rocks found in the study region (GSI 2006). Clayey soil, fine loamy mixed soil and clayey-skeletal kaolinite soils are found in the studied region (NBSS 1996). Geomorphologically, the study region from east to west may be divided into three primary sections: the high escarpment of the Western Ghat, the mild plateau and the deep valley. Idukki District has elevation values which range from 15 to 2700 m above sea level. The research area is frequently flooded near the bottom of the Ghat region due to strong orographic rainfall during the monsoon season. Many tiny communities are located in such areas, and they are severely impacted both physically and economically.

The Idukki district is renowned for its hilly terrain and diverse range of flora and fauna. During August of 2018, Kerala experienced a catastrophic flood event, with the Idukki district being among the areas that were severely affected. In August 2018,

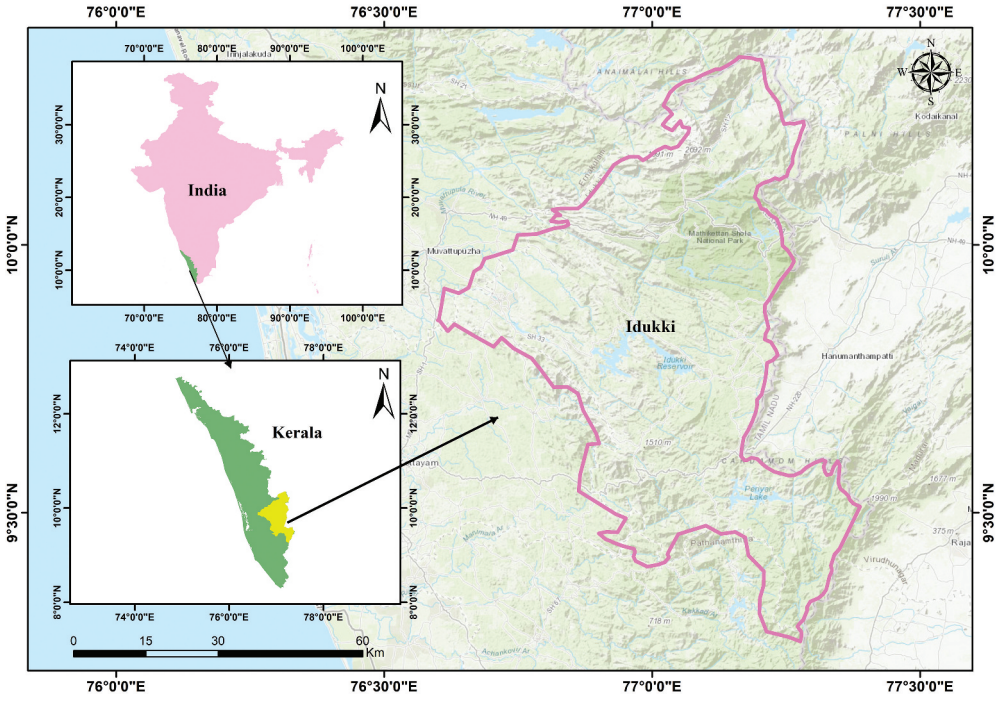


Figure 1. Location map of the study area.

a period of prolonged flooding was observed over multiple days. The precipitation of significant magnitude and consequent inundation were instigated by the southwest monsoon. The region experienced a significant amount of precipitation during the monsoon season, with approximately half of the total rainfall occurring within the initial 20-day period of August 2018. This resulted in a swift and excessive inundation of the area. Numerous regions located in the district encountered large quantities of precipitation, leading to a notable escalation in the water levels of rivers, lakes and reservoirs. As a result of the significant precipitation and the imperative to regulate the escalating water levels in the reservoir, the flood gates of the Idukki dam were opened, thereby contributing to the inundation of the lower-lying areas. The hydrological phenomenon of dam water release was a contributing factor to the extensive inundation observed in the district. The Idukki district suffered significant infrastructure damage as a result of the floods. The occurrence of landslides in mountainous regions exacerbated the situation by causing interruptions in transportation and communication infrastructures. The inundation that occurred in the Idukki district in 2018 led to a noteworthy fatality rate. Numerous individuals were left stranded and compelled to evacuate their residences as a result of the escalating floodwaters. The establishment of relief camps aimed to offer lodging, sustenance and healthcare aid to the impacted populace. The inundations had a significant effect on the agricultural sector and the means of subsistence in the region. The inundation of crops, encompassing paddy fields, plantations and horticultural lands, led to significant economic losses for agricultural producers. The livestock and fisheries sectors were also impacted.

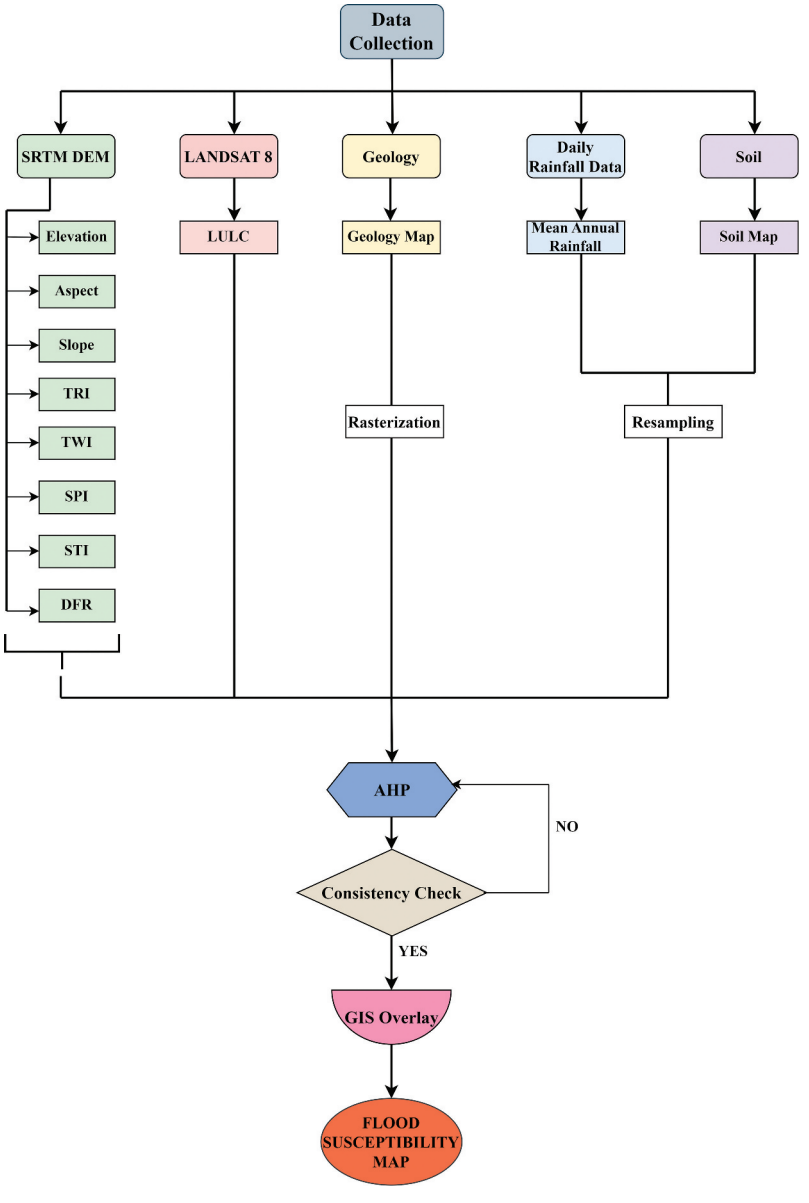


Figure 2. Methodology for preparing flood susceptibility map of Idukki district.

3. Material and methods

3.1. Methods for preparing flood thematic layers

The creation of the spatial flood model, which is frequently highly intricate and far from exact, necessitates a large amount of topographical and hydrological data. The flood susceptibility maps will be confirmed using scientifically defined metrics. As a result, one of the most important tasks is to identify the variables that cause flooding. The 12 flood-inducing parameters in the current study area were chosen based on the available flood

susceptibility literature which includes geology, elevation, distance from river, rainfall, land use, soil, topographic wetness index (TWI), slope, topographic roughness index (TRI), aspect, stream power index (SPI) and sediment transport index (STI) (Moghaddam *et al*. 2019, Paul *et al*. 2019, Sturzenegger *et al*. 2019). All of the influencing variables were transformed to a raster format with a spatial resolution of 30 metres by the resample tool in ArcGIS.

For flood modelling, topographic parameters are critical because they influence the hydrological properties of the studied region both directly and indirectly (Arabameri *et al*. 2020, Lei *et al*. 2020). Das and Gupta (2021) developed a flood susceptibility map for Subarnarekha River basin, India, and indicated that TWI, elevation, drainage density and slope are the most crucial factors for flood hazard mapping. In the first phase, the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) for the research region was downloaded from Earth Explorer. This consists of 1 arc second 30 m resolution freely available DEM data widely used in various applications of remote sensing analysis. These data are re-projected to UTM zone 43 N in the projected coordinate system. The data are void filled, mosaicked and extracted by mask for further processing in the ArcGIS 10.2 environment. These pre-processed SRTM DEMs can provide more accurate and reliable elevation data for flood modelling. We used DEM to compute slope, aspect, curvature, sediment transport index (STI), topographic wetness index (TWI), stream power index (SPI) and topographic roughness index (TRI) in the ArcGIS environment. Each of the thematic layers were obtained from different sources as shown in the Table 1. The methodology for preparing flood susceptibility map of Idukki district is illustrated in Figure 2.

The most important aspect in flood modelling is elevation (Dodangeh *et al*. 2020). Flooding is inversely proportional to elevation. When heavy rainfall occurs, water flows quickly from higher-elevation regions such as mountainous and hilly areas to low-lying areas or flat regions of lower elevation, thereby making these low-lying areas more prone to flooding. (Chen *et al*. 2019, Veerappan and Sayed 2020). As a result, floods are more likely to occur on flat or lower-elevated areas. Figure 3(a) shows the elevation map of the study region. Another factor that influences the direction of flooded areas is aspect. This is derived from the DEM data spatial analyst tool in ArcGIS. It is considered to be a significant factor as it determines the direction of flow of water and hence assists in the evaluation of flooded regions (Rehman *et al*. 2021). As a result, the aspect has an indirect impact on floods. Figure 3(b) shows the aspect of the study area. The slope, which affects the flowing water's

Table 1. Database information of flood susceptibility mapping.

Data layers	Source of data	Data type	Resolution/ feature type
Rainfall	Indian Meteorological Department https://mausam.imd.gov.in/	Gridded	0.25°
Geology/lithology	Geological Survey of India https://www.gsi.gov.in	Vector	Polygons
Soil	National Bureau of Soil Survey and Land Use Planning https://www.nbsslup.in/	Raster	2.5 km
Digital elevation model (DEM)	SRTM https://earthexplorer.usgs.gov/	Raster	30 m
Land-use/land-cover	Landsat 8 OLI Image https://earthexplorer.usgs.gov/	Raster	30 m

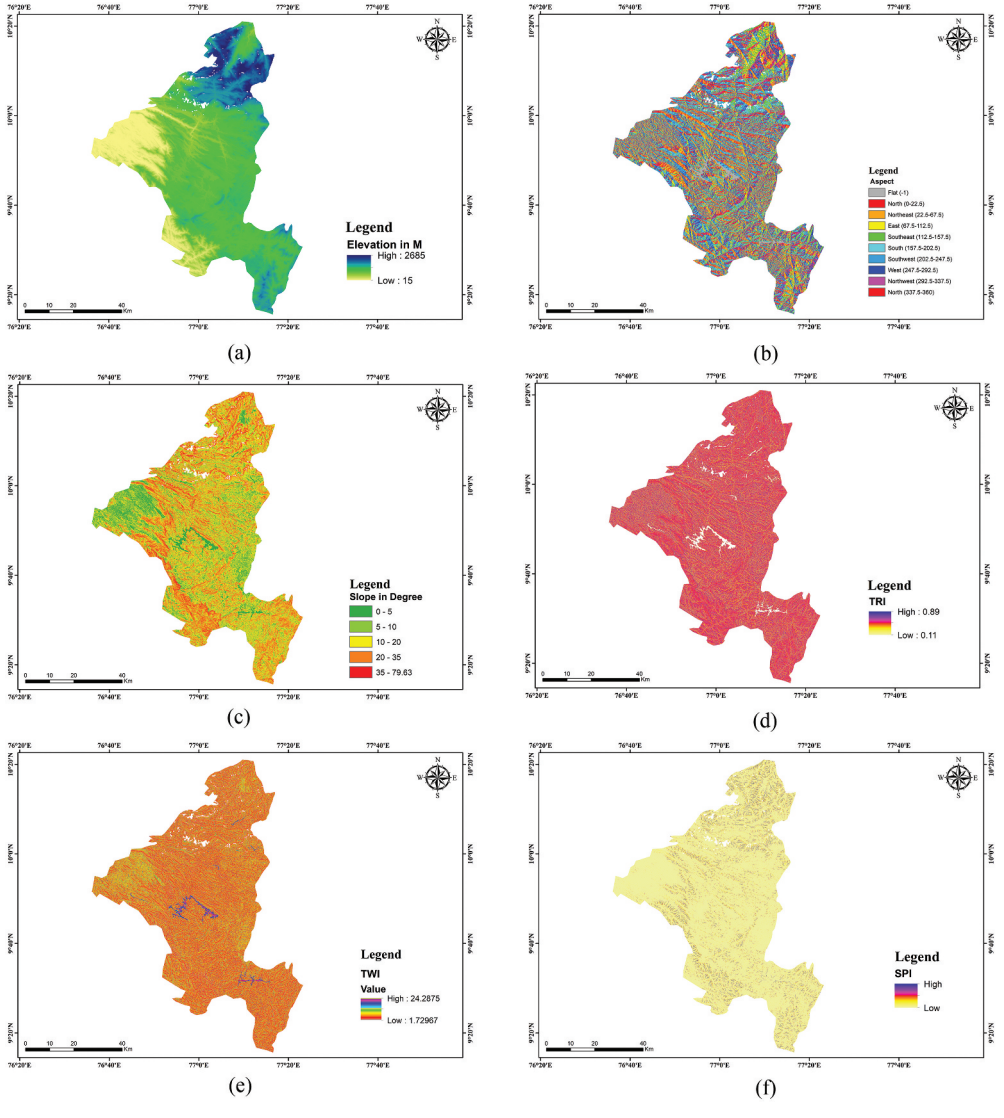


Figure 3. Thematic layers: (a) elevation, (b) aspect, (c) slope, (d) topographic roughness index, (e) topographic wetness index, (f) stream power index.

flooding pace, is another major aspect that influences the flood (Stevaux *et al*. 2020). Slope is derived from the DEM data using the spatial analyst tool in ArcGIS. In regions where the slope is steep, water flows with a high velocity during a rainfall event and hence disposes the surface runoff very quickly, whereas areas of flatter or gentle slopes are more prone to floods and water logging, as runoff gets stored over these regions for a longer time. Figure 3(c) demonstrates the slope of the study area. TRI is also a key contributing variable for flood episodes as it assists in investigating basin's local topography. It is calculated by deriving the focal mean, focal minimum and focal maximum of the raster in each pixel in the ArcGIS environment and using the following equation in the raster calculator.

$$TRI = \frac{(Focal\ mean - Focal\ minimum)}{(Focal\ maximum - Focal\ minimum)}$$

It describes the amount of difference in elevation between adjacent cells in a DEM. Higher magnitude of floods is always related to lower TRI values and a high value of TRI leads to low magnitude of floods (Talukdar *et al*. 2020). The TRI (topographic roughness index) map was preserved in stretch format in this investigation, with values ranging from 0.11 to 0.89 (Figure 3(d)). TWI (topographic wetness index) is also an important aspect in the occurrence of floods (Abdel Hamid *et al*. 2020), since it allows the difference in the wetness of a basin to be visualised spatially (Meles *et al*. 2020). This index depicts the water quantity present in each pixel of the area, and it is calculated using the following equation in the raster calculator in the ArcGIS environment.

$$TWI = \frac{ln(A_s)}{tan\beta}$$

A_s indicates the catchment area ($m^2\ m^{-1}$) and β represents local slope gradient (in degrees). Floods and high TWI levels are inextricably related. In the study region, the TWI value ranges from 1.72 to 24.28 (Figure 3(e)).

The stream power index (SPI) has a significant influence on the riverine system. A higher value of SPI is indicative of large erosion and heavy flowing water. SPI is calculated using the equation in the raster calculator in the ArcGIS environment.

$$SPI = A_s\ tan\beta$$

The specific catchment area is represented by A_s ($m^2\ m^{-1}$) and the local slope gradient is denoted by β (in degrees) (Wu *et al*. 2020). The SPI map is shown in Figure 3(f). The sediment transport index (STI) describes the topographical effect on loss of soil and erosive capacity of a terrain. It is derived from DEM using the following equation in the raster calculator in the ArcGIS environment.

$$STI = \left(\frac{A_s}{22.13}\right)^{0.6}\left(\frac{Sin\beta}{0.0896}\right)^{1.3}$$

where β denotes each pixel of the local slope (in degrees) and the area of the upstream region is denoted by A_s ($m^2\ m^{-1}$). The STI value ranges in this study were classed from low to high (Figure 4(a)).

Surface runoff and sediment transport are influenced by land use, which has a direct impact on the frequency of floods (Benito *et al*. 2010). Land use has a direct influence on surface runoff generation and infiltration, because metropolitan areas do not permit water to enter the ground, as a result of which flooding is more prevalent. In this work, we used the ArcGIS image classification approach to create a land use map from a Landsat 8 OLI image. The land use map was divided into seven categories: barren and fallow land, cropland, forest, grassland, plantation, settlements and waterbodies (Figure 4(b)). Supervised classification was employed to categorise land use. Supervised classification allows a user to select sample pixels in an image that are representative of specific classes and directs the image processing software to utilise these training samples as a reference for the classification of all other pixels in the image. Training samples are chosen based on the knowledge and expertise of the user. For instance, training areas for urban areas, vegetation and water bodies can be selected by drawing

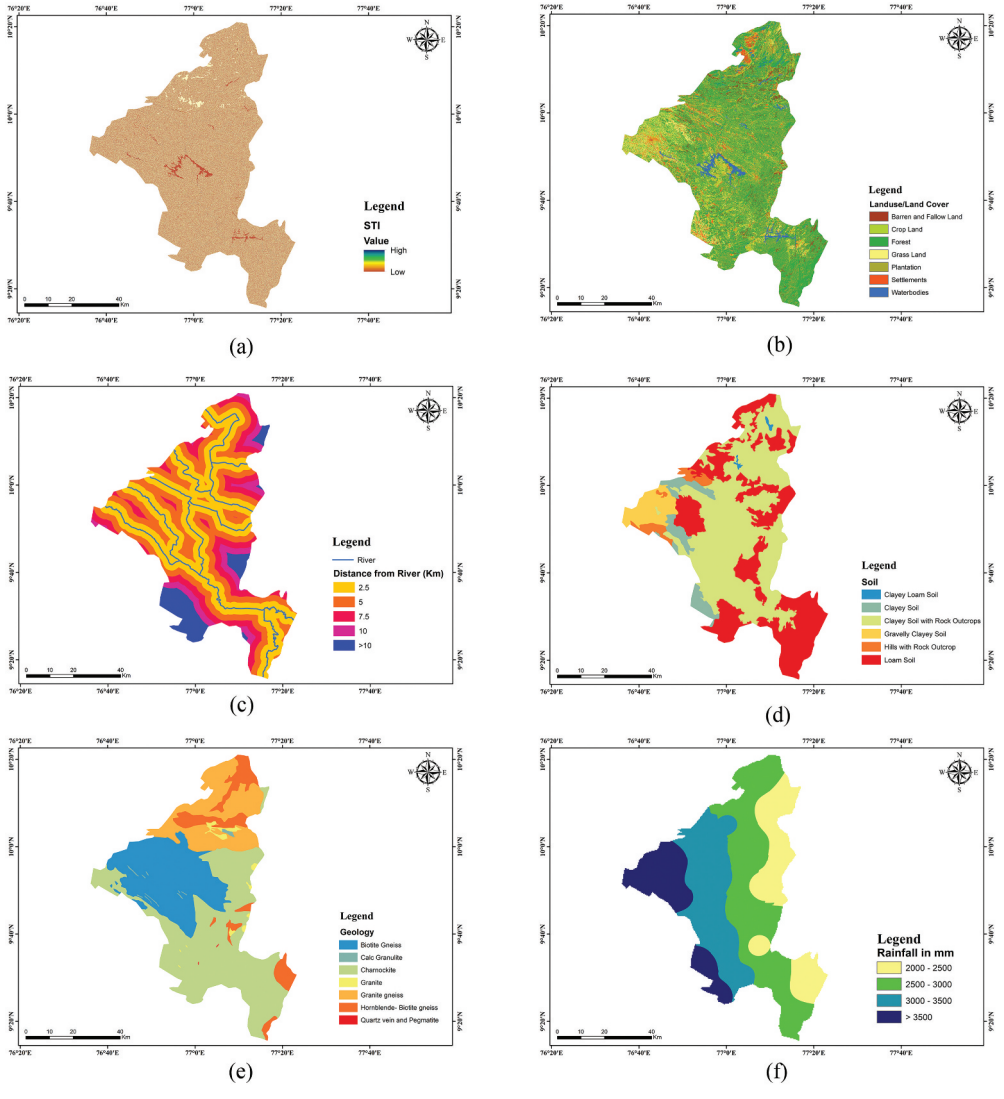


Figure 4. Thematic layers: (a) sediment transport index, (b) land use, (c) distance from river (DFR), (d) soil, (e) geology, (f) rainfall.

a polygon around them. After producing enough training samples for each class, signature files are generated in the ArcGIS software. Using the supervised classification tool, the study region is then classified into different land use features. The precision of the classification process is determined by selecting spatially balanced points within the research area and cross-referencing them with manually verified ground truth data obtained through the use of Google Earth Pro software. Subsequently, a confusion matrix was generated to evaluate the precision and reliability of the classified map. The results indicate an accuracy rate of approximately 83.35%, which is considered acceptable. Most flood-inundation areas are often located in the vicinity of a river or a stream. Distance from the river is a critical parameter for defining the areas of a basin that are prone to flooding because it governs the incidence of flooding and the river flow to river factor. The underlying concept is that the less the distance from the river,

the more likely it is that the region will get flooded. In the current study, we computed the distance to the river map using a 1:50,000 scale topographic map and Google Earth. The distance from the river chart is depicted in Figure 4(c). Soil types are one of the most important influencing elements in the rainfall-runoff process. Other variables such as local meteorological conditions and erosion regulate rainfall-runoff production, but soil qualities directly limit water infiltration, which determines rainfall-runoff generation. In the present study, we have used the National Bureau of Soil Survey and Land Use Planning (NBSS and LS) to prepare the soil map. Figure 4(d) shows the soil map of the study area. The lithological properties of a region influence porosity and permeability. Properties of a particular rock can also impact flood intensity to some extent because in the presence of impermeable rocks more flooding is bound to occur, while porous rocks will let more water infiltrate and hence lessen the chances of flooding. In the present study, the geology map was prepared using Geological Survey of India data (1:250,000 scale). Most of the study area is covered by charnockite followed by biotite gneiss, granite gneiss, etc (Figure 4(e)). Rainfall has been identified as one of the most important elements determining the incidence of floods because a large amount of rain in a short period of time might produce floods. Annual average rainfall data were collected from 15 Indian meteorological stations and with the help of the inverse distance weighted (IDW) interpolation technique, rainfall maps were created in the ArcGIS environment. IDW estimates the values of cells by averaging the values of sample data points in the neighbourhood of each processing cell. The closer a point is to the centre of the cell being estimated, the more influence it has in the averaging process. The disadvantage associated with IDW interpolation is that it is not suitable for territories having large orographic effects; however, due to the distribution of our rainfall data, this technique produced accurate rainfall maps. The annual rainfall ranges from 2000 to 3900 mm in the study area as shown in Figure 4(f).

3.2. Analytical hierarchy process (AHP) model

The analytical hierarchy process method is a versatile and well-structured technique for interpreting complex decision scenarios involving several factors (Saaty 1980). It is a well-established approach for solving complex decision-making problems. The advantage of using this method lies in the allotment of user-specified weights to a hierarchical structure of criteria and sub-criteria. Therefore, AHP assists the user to evaluate different parameters that are of utmost significance as per the requirement and to comprehend the problem statement with deeper understanding. Comparison of different parameters is done by assigning scores based on their relative influences and importance (i.e. 1 denotes equally significant; 3 denotes moderately more significant; 5 denotes strongly more significant; 7 denotes very strong more significant; 2, 4 and 6 are the transition scores) as shown in Table 2. In various problems such as flood susceptible mapping, landslide mapping and groundwater potential zone assessment, AHP is found to be a powerful and a beneficial tool having high accuracy (Mahmoud and Gan 2018, Surwase *et al*. 2019, Suthirat *et al*. 2020). The reason for the wide popularity of the AHP model is its capability to handle quantitative as well as qualitative data in assessing natural hazards like floods. Implementation of the AHP model has been successful in comprehending problems of various natural hazards such as assessment of flood and landslide susceptible regions (Danumah *et al*. 2016, El Jazouli *et al*. 2019).

Table 2. Saaty's 1987 1–9 scale of pairwise comparison matrix.

Numerical value	Description	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Slight importance of one over another	Experience and judgement slightly favour one activity over another
5	Moderate importance of one over another	Experience and judgement strongly favour one activity over another
7	Very strong importance	Activity is favoured very strongly over other
9	Extreme importance of one over another	Evidence favouring one activity over the other is of the highest importance
2, 4, 6, 8	Intermediate values	-

For investigating problems of flood modelling or susceptibility studies, the AHP model is used to estimate the weights and rankings of various factors. Steps involved in the AHP method for delineation of flood susceptibility zones in a region are described below:

- The first step of AHP is preparation of a pairwise comparison matrix of all the parameters affecting a system.
- The weights or values then are assigned to each parameter purely based on the knowledge or the expertise of the user in the pairwise comparison matrix. For instance, when rainfall is compared with elevation, a value of 4 is assigned which indicates intensity of importance for rainfall is 4 times more than that of elevation, as shown in Table 3.
- Similarly, when the same parameters are compared in the standard pairwise comparison matrix, for instance when slope is compared with slope, the assigned weight has a value equal to 1 as both the parameters hold equal importance, as shown in Table 3. This further explains the reason for assigning a value of 1 to all the diagonal elements in the pairwise comparison matrix in Table 3. Evaluation of all the attributes is generally performed on a 9-point scale, as explained by Saaty (1980), in Table 2.
- The second step is to sum all the assigned weights at the end of a column and then each assigned weight is divided by sum of all the assigned weights. The normalised principal eigenvector or priority vector is obtained by averaging all the weights and

Table 3. Each parameter's comparison matrix and relative score.

Parameters	Rainfall	Elevation	DFR	TWI	Slope	TRI	SPI	STI	LULC	Geology	Aspect	Soil
Rainfall	1.00	4.00	5.00	4.00	3.00	3.00	4.50	6.50	7.00	6.50	8.50	8.50
Elevation	0.25	1.00	3.00	2.00	3.00	2.50	3.00	4.00	5.00	4.00	7.50	7.00
DFR	0.20	0.33	1.00	5.00	4.50	3.00	2.00	3.50	5.00	3.50	5.50	7.00
TWI	0.25	0.50	0.20	1.00	4.00	2.00	3.00	3.00	3.50	4.50	4.00	5.50
Slope	0.33	0.33	0.22	0.25	1.00	4.00	3.50	4.00	3.50	3.50	4.50	5.00
TRI	0.33	0.40	0.33	0.50	0.25	1.00	2.00	3.00	2.00	2.50	4.00	3.00
SPI	0.22	0.33	0.50	0.33	0.25	0.50	1.00	5.00	2.50	2.00	3.00	3.50
STI	0.15	0.25	0.29	0.33	0.25	0.33	0.20	1.00	3.00	3.00	2.00	2.50
LULC	0.14	0.20	0.20	0.29	0.29	0.50	0.40	0.33	1.00	2.00	2.00	4.00
Geology	0.15	0.25	0.29	0.22	0.29	0.40	0.50	0.33	0.50	1.00	2.00	3.00
Aspect	0.12	0.13	0.18	0.25	0.22	0.25	0.33	0.50	0.50	0.50	1.00	2.00
Soil	0.12	0.14	0.14	0.18	0.20	0.33	0.29	0.40	0.25	0.33	0.50	1.00
Sum	3.27	7.88	11.35	14.36	17.24	17.82	20.72	31.57	33.75	33.33	44.50	52.00

a normalised value is obtained for each parameter. The priority vector indicates the final weights assigned to each parameter.
- To evaluate the consistency of weights with respect to one another, the consistency ratio is calculated.
- After the consistency check, the normalised weightage is then multiplied with rank of all individual parameter sub-classes to get the overall weightage of all thematic layers (Table 4).
- Finally, all the weightage values are attributed in the ArcGIS file to get the final flood susceptibility map.

A detailed mathematical expression of the AHP approach is given below.

Let X = {X_j| j = 1, 2 . . . . . n} be the arrangement of the model. The result of the pairwise comparison on n criteria can be done in a (n_n) assessment matrix A where each component a_{ij} (i, j = 1, 2 . . . n) is the range of the weight of the attribute, as given in the condition below

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \; a_{ii} = 1, \; a_{ji} = \frac{1}{a_{ij}}, \; a_{ij} \neq 0$$

The weights or scores of pairwise comparison matrix are normalised to evaluate the standard pairwise comparison matrix through the method of weighted arithmetic mean. The weights of all the components are determined by the mean row method in the standard pairwise comparison matrix. The maximum characteristic root can be expressed as:

$A\ w = \lambda\max\ w$. Here A represents the standard pairwise comparison matrix, w is the eigenvector, λmax represents the matrix's principal eigenvalue.

To have a consistent pairwise comparison, the matrix A has rank equivalent to 1 and λmax is equivalent to n. Consistency of the pairwise judgements among the different parameters is of utmost importance as it regulates the accuracy of the results through the AHP model. The consistency of the AHP method is checked by employing the equation below (Saaty 1980):

$$CR = \frac{CI}{RI}$$

$$CI = \frac{\lambda\max - \mathrm{n}}{n - 1}$$

where CR represents consistency ratio, CI represents consistency index, RI represents random index, λmax represents the matrix's principal eigenvalue, and n denotes the number of parameters or factors. In Saaty's (1980) paper, a significant number of components are used to obtain RI values, as shown in Table 5. It has been discovered that RI is influenced by a variety of circumstances. According to Saaty's (1980) study, for instance, the comparable value of RI is 1.45 when the components are nine (Table 5). The consistency ratio (CR) is acceptable in this study if it is less than 10% or 0.10, which suggests a fair level for the pairwise comparison matrix.

Table 4. Sub-criteria of each parameter and their weights.

Parameter	Factor	Weight	Rank	Overall weight
Elevation	15–250	0.15	1	0.15
	250–500		2	0.30
	500–1000		3	0.45
	1000–1500		4	0.60
	>1500		5	0.75
Slope	0 to 5	0.10	5	0.50
	05 to 10		4	0.40
	10 to 20		3	0.30
	20 to 35		2	0.20
	>35		1	0.10
Distance from river	2.5	0.14	5	0.70
	5		4	0.64
	7.5		3	0.42
	10		2	0.28
	>10		1	0.14
LULC	Barren and fallow land	0.03	4	0.12
	Crop land		3	0.09
	Forest		4	0.12
	Grass land		3	0.09
	Plantation		3	0.09
	Settlements		1	0.03
	Water bodies		5	0.15
Geology	Biotite gneiss	0.03	2	0.06
	Calc granulite		3	0.09
	Charnockite		5	0.15
	Granite		5	0.15
	Granite gneiss		4	0.12
	Hornblende biotite gneiss		2	0.06
	Quartz vein and pegmatite		4	0.12
Soil	Clayey loam soil	0.02	4	0.08
	Clayey soil		3	0.06
	Clayey soil with rock outcrops		5	0.10
	Gravelly clayey soil		4	0.08
	Hills with rock outcrop		5	0.10
	Loam soil		2	0.04
TWI	1.31–1.88	0.10	1	0.01
	1.89–3.51		2	0.02
	3.52–5.96		3	0.03
	5.97–9.64		4	0.04
	9.65–19.44		5	0.05
Rainfall	2000–2500	0.25	1	0.25
	2500–3000		2	0.50
	3000–3500		4	0.10
	>3500		5	0.12
TRI	0.1–0.45	0.06	3	0.18
	0.45–0.52		4	0.24
	0.52–0.88		5	0.30
Aspect	Flat	0.03	3	0.09
	North		4	0.12
	Northeast		2	0.06
	East		1	0.03
	Southeast		2	0.06
	South		4	0.12
	Southwest		5	0.15
	West		5	0.15
	Northwest		5	0.15
	North		4	0.12
SPI	−7.42 – −5.36	0.06	1	0.06
	−5.36 – −2.46		2	0.12
	−2.46 – −0.62		3	0.18
	−0.62–2.19		4	0.24
	2.19 – 12.04		5	0.30
STI	High	0.04	5	0.20
	Moderate		4	0.16
	Low		3	0.12

Table 5. Saaty's random index (*RI*) table.

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.48	1.56	1.57	1.58

Generally, for flood-related studies, previous research scholars generally utilise eight or nine parameters (Das 2018, Mandal and Mandal 2018, Saha and Agrawal 2020). In Idukki District, previous studies related to landslide susceptibility mapping have often been performed (Thomas *et al*. 2021, Jesudasan and Subbarayan 2022) to assess the risk-prone regions related to this natural hazard. However, studies related to floods have not yet been performed in Idukki District, and since it is one of worst affected districts in Kerala floods every year, it becomes essential to map the flood-prone areas of the district with accuracy. Hence, the present study has considered as many as 12 flood conditioning factors, which are elevation, slope, rainfall, topographic wetness index, topographic roughness index, stream power index, distance from the river, sediment transport index, land use, geology, aspect and soil for accurately identifying flood-prone zones using the AHP model in Idukki District. In earlier studies, only primary terrain attributes such as aspect, slope, elevation, soil types and geology were the main conditioning parameters in computing flood-prone areas. However, secondary terrain attributes or topo-hydrological aspects such as TWI, TRI, STI and SPI are also of paramount significance as they determine the effect of local topography, detailed undulations, overland flow and erosion risks. Hence the current study also includes the attributes of topo-hydrological parameters along with primary terrain attributes for calculation of flood-prone zones with utmost accuracy.

From consistency table and calculations, λmax denotes the principal eigenvalue, *n* refers to the dimension of the matrix and calculation of λmax and Consistency Index is given below:

$$\lambda\text{max} = 13.64416551$$

$$\text{Consistency Index} = \frac{\lambda\,\text{max} - 1}{n - 1}$$

$\text{Consistency Ratio} = \frac{\textit{Consistency Index}}{\textit{Random Index}} = 0.09706$, Here, the value of the Random Index (*RI*) is taken as 1.48 as there are 12 flood conditioning factors. Consistency ratio for the current study is less than 0.10 or 10%, hence weightages given to all the parameters are consistent with one another. This assists in having a preliminary quality assessment for the final map, as the resultant map is based on the assigned weights of the different parameters, hence inaccuracy in assigning weights would create bias in the results. It also implies that by employing these weights for the development of the flood susceptibility map, accurate assessment of the flood-prone areas can be performed for the Idukki district as all the flood conditioning factors are consistent with each other.

4. Results

All the thematic layers of 12 flood conditioning factors are prepared in the GIS interface as shown in Figures 3 and 4. After preparation of all the thematic layers, relative weights are given to all the flood conditioning parameters and a comparison matrix is made as shown in Table 3. The consistent weights are then calculated for every parameter in the standard pairwise comparison matrix. Since a lot of parameters are utilised for preparing a flood susceptibility map, a consistency check is of utmost importance to ensure that weightages given to parameters are consistent with each other.

The consistency check for all the flood conditioning parameters is evaluated, where the consistency ratio is found to be 0.09706. For maintaining consistency between parameters, the consistency ratio must be less than 10%, and hence the weightage given to each flood conditioning parameter is consistent with each other. AHP weightages for all the thematic layers, namely rainfall, elevation, distance from river, TWI, slope, TRI, SPI, STI, LULC, geology, aspect and soil ares 0.25, 0.15, 0.14, 0.10, 0.10, 0.06, 0.06, 0.04, 0.03, 0.03, 0.03 and 0.02 respectively. These weightage values are further multiplied with all the classes and ranks to determine the effect of each parameter and a final flood susceptibility map is developed, as shown in Figure 5.

The final flood susceptibility classes are ranged from 0.25 to 0.02 and are classified using natural break classification into five categories namely very low, low, moderate, high, and very high and these classes occupied areas of 609.0417 km^2, 1222.83 km^2, 1180.45 km^2, 950.48 km^2 and 395.9487 km^2 respectively (Table 6). More than 30% of the Idukki region is covered with high and very high flood susceptible zones (Figure 5). These very high and high flood risk zones were derived from the areas of low elevation, low slope area (less than 10°), distance from the river less than 2.5 km, highest topographic wetness index region (>15), very high rainfall region (3000–3500 mm), highest stream power and sediment transport index region and finally hard rock terrain and low permeability soil type region. Figure 5 displays the flood susceptibility map of the current study area. The very high and high flood-prone regions are in NW, central and SE regions of the study area.

From the flood susceptibility map shown in Figure 5, it is worthwhile to note that more than 30% of the area (1346.4 km^2) falls in the category of very high to high flood-prone zones (Table 6). Therefore, it becomes very important to assess the impact of these very high flood risk zones at subdivisional levels of Idukki. There are five taluks or subdivisions in Idukki, namely Devikulam, Udumbanchola, Idukki, Peerumade and Thodupuzha, as marked in Figure 6, which categorises the flood prone zones for every taluk separately. The developed flood susceptibility map indicates that majority of Thodupuzha town falls in very high flood-prone zones. This is because Thodupuzha lies in areas of lower elevation, low slope and high rainfall as seen from Figures 3(a,c) and 4(f). The distance from the river as shown in Figure 4(c) also reveals that Thodupuzha and Idukki have dense stream networks and major river flow through these districts, which become the paramount reason for having a high likelihood of flooding in these towns. During the Kerala floods of 2018, among all the towns Thodupuzha and Idukki were the worst affected. Seventeen villages constitute the Thodupuzha taluk or subdivision, which shows that people in these villages are at a very high risk of flooding and substantial damage due to floods in future could occur in these areas. Furthermore, the majority of the Idukki subdivision falls under high to very high flood-prone zones. This is because of the presence of the Idukki dam and a large number of

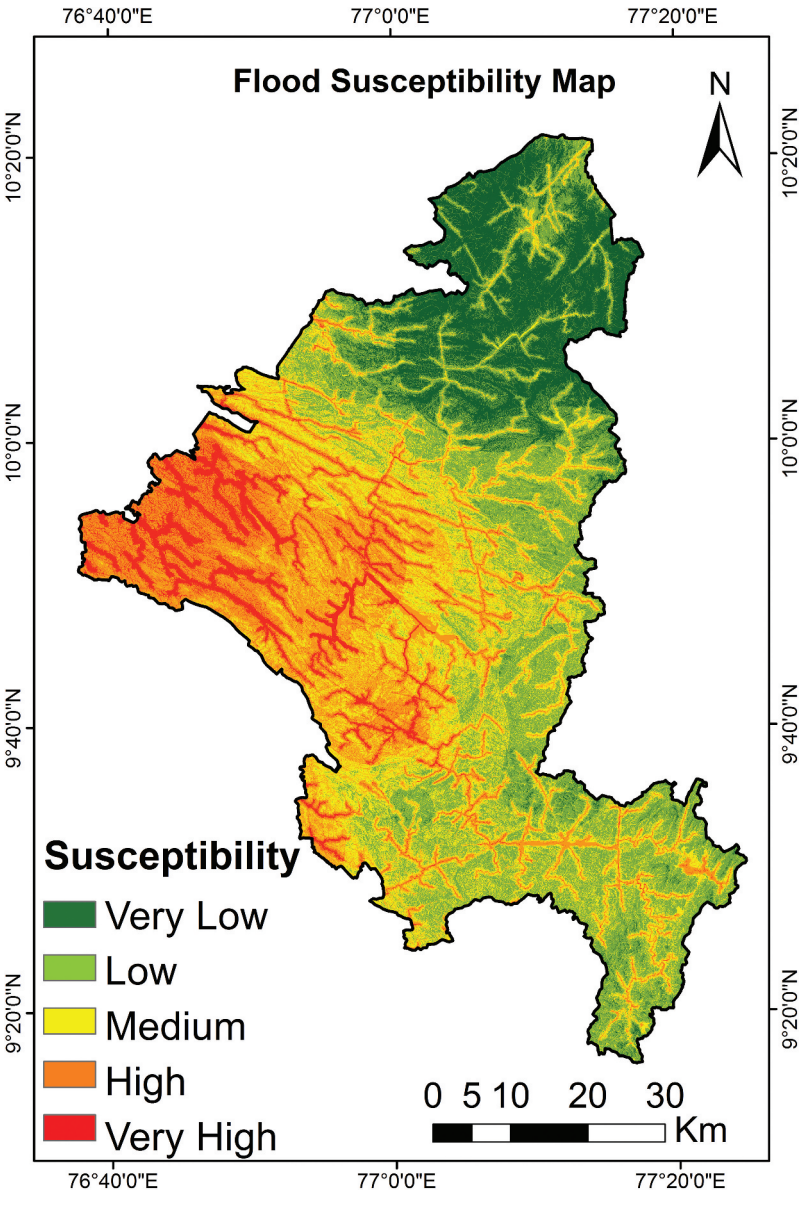


Figure 5. Resultant flood susceptibility map of Idukki district.

Table 6. Flood-prone zones.

S. No.	Flood-prone zones	Area (km^2)
1.	Very low	609.0417
2.	Low	1222.8336
3.	Moderate	1180.4571
4.	High	950.4855
5.	Very high	395.9487

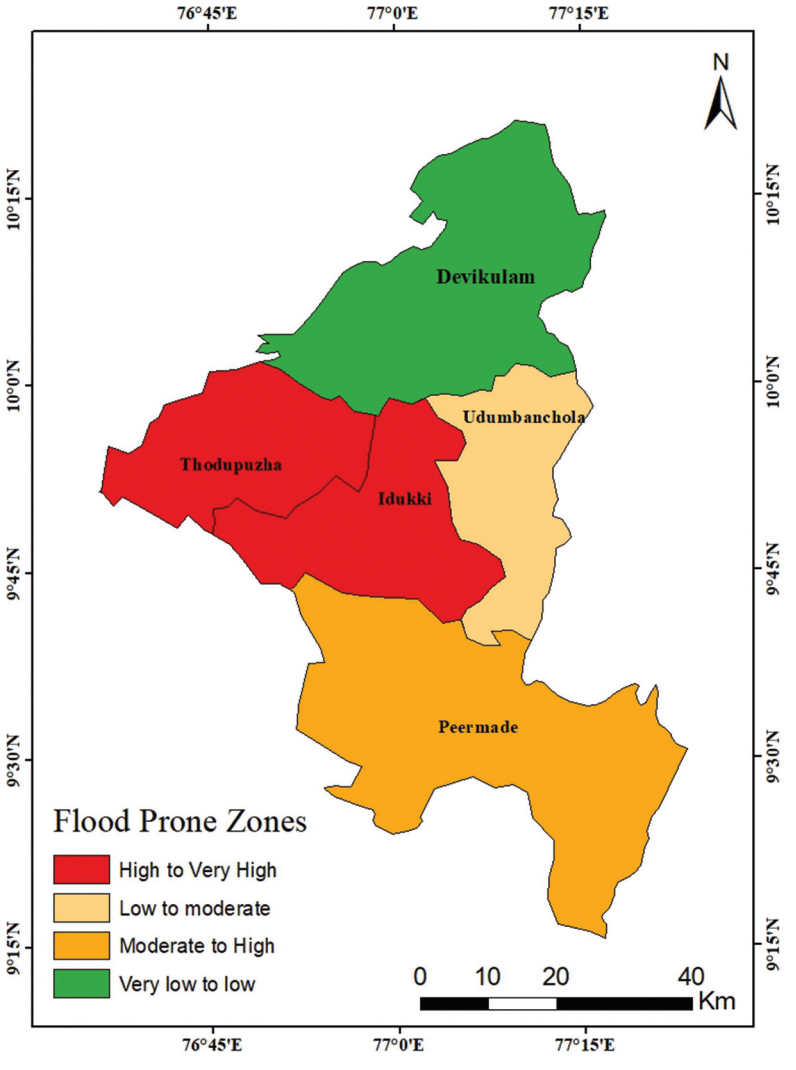


Figure 6. Categories of flood-prone zones in subdivisions of Idukki.

tributaries of the Periyar river. During the Kerala floods of August 2018, all five gates of the dam were opened by the authorities after 26 years, which caused a massive outflow of six lakh litres per second along the Periyar river and its tributaries. Due to this reason, the Periyar river and its tributaries flowed under spate conditions, which leads to flooding of the river banks causing damage to the people residing in these areas. Nine villages constitute the Idukki town and are home to many people, therefore inhabitants residing in these villages are at very high risk of flood-related natural hazards. Peermade subdivision, which is situated in the south east region of Idukki, falls under moderate to high flood prone zones, as shown in the developed flood susceptibility map in Figures 5 and 6. This region generally receives a considerable amount of rainfall in the monsoon season. During the Kerala floods of 2018 and 2021, this region received rainfall in the range 300–350 mm per day and some portion falls under low-lying areas, therefore the majority of the region falls under moderate

to high zones of flooding. Udumbanchola falls under low to moderate flood-prone zones as the region has a dense stream network, high TWI values and low rainfall. From the flood susceptibility map, it is clearly depicted that Devikulam taluk falls under low to very low flood-prone zones because of higher elevation values, a smaller number of streams, low TWI values and comparatively less rainfall than other subdivisions. The flood-prone zones in various subdivisions of Idukki, along with the taluks of maximum flood-prone regions and least flood-prone areas are shown in Figure 6. The process involves partitioning the vast majority of susceptible areas into distinct subcategories based on taluks. This measure is implemented with the aim of decreasing the affected population in that particular locality in the event of a disaster. The detailed impact of all 12 flood conditioning factors on the flood susceptibility map are also discussed below.

4.1. Impact of elevation

This largely influences floods in a region as low-lying areas are most likely to be affected by floods. As shown in the flood susceptibility map (Figure 5), it clearly indicates that subdivisions such as Thodupuzha, Idukki and Udumbanchola which are situated on comparatively lower elevation come under high- to very high-risk zones in flood susceptibility mapping (Xu *et al*. 2021).

4.2. Impact of slope

Another factor which largely influences and controls the flood is slope. In regions of higher slope, runoff flows with a high velocity and chances of water infiltrating through soil become much less and hence water stagnates in areas with small or flat slopes. This is evident as western and central regions of Idukki have low slope and hence are categorised in the areas of very high flood susceptibility (Figure 5). Though regions in the north of Idukki have comparatively higher slope, yet we still see medium flood risk in the flood susceptibility map (Figure 5). This is due to the presence of a large number of streams such as Pembar, Edamalaya and Muthirapuzha rivers, which overflow in events of heavy rainfall and hence make the region prone to flooding Hunt & Menon (2020).

4.3. Impact of TRI

Higher TRI indicates steep and rugged regions, whereas lower TRI indicates flatter areas. As Idukki is situated in a mountainous area factor, such as TRI in conjunction with slope assist in accurate assessment of flood mapping.

4.4. Impact of TWI

Topographic wetness index is an indicator of depth of surface and subsoil layers. Higher TWI indicates the presence of larger streams while lower values give an idea of ridge lines (Pourali *et al*. 2016). In the flood susceptibility map, higher values of TWI can be seen in the subdivisions of Thodupuzha and Peerumade and hence these regions fall in the category of high to very high flood-prone zones.

4.5. Impact of SPI

SPI is an indicator of the erosive power of surface runoff. It is a measure of flowing water's ability to perform a geomorphic task. Geomorphic tasks in relation to flood susceptibility studies would involve the occurrence of weathering, deposition and erosion processes over a land surface during events of heavy rainfall and flooding. Higher values of SPI can be seen in the northern regions of Idukki, which implies large erosion and heavy flowing water. This is because of mountainous regions, where water flows with a high velocity and causes erosion near the river bank. The resulting flood susceptibility map (Figure 5) reveals that low- to medium-risk flood zones are present in this area.

4.6. Impact of STI

Sediment transport index characterises the process of erosion and deposition. It is indicative of the effect of topography on soil loss. Higher STI values represent higher mobility of sediments while the lower values indicate the accumulation of sediments. It is used to describe the erosive capacity of a terrain and the possibility of changes in the channel bed geometry due to sediment deposition. Higher values of STI are found in the streams present in the Idukki district due to heavy flowing water, which helps in sediment deposition near the streams. Moreover, the flood susceptibility map also depicts very high risk of flooding in or near the streams. Therefore, STI develops a direct relation with the flood risk map. Lower STI values are present in the Idukki reservoir and Periyar lake, which is in sync with the flood susceptibility map. (Nones 2019, Siam *et al*. 2022)

4.7. Impact of LULC

LULC has identified seven major classes to quantify the impact of land use and land cover on floods. The capability of land cover types to either absorb rainfall or facilitate its infiltration into the soil varies. The presence of natural vegetation, such as forests or grasslands, has the potential to intercept rainfall and promote infiltration, thereby mitigating the volume of surface runoff. Conversely, urban regions or impermeable surfaces exhibit restricted infiltration potential, thereby leading to escalated surface runoff and expedited water flow towards streams and rivers, thereby augmenting the likelihood of inundation. Rough terrain, such as that found in wooded areas or open grasslands, can impede the velocity of water flow, thereby facilitating its infiltration or retention within the topography. Conversely, surfaces that are devoid of roughness such as roads tend to accelerate the flow of water, resulting in increased quantities of runoff and the possibility of inundation. The cutting down of forests or vegetation cover results in the disturbance of the natural equilibrium of water absorption and interception. The process of deforestation has been observed to have a negative impact on the hydrological cycle, as it leads to a reduction in the amount of rainfall that is intercepted by trees. Additionally, deforestation has been found to increase surface runoff and decrease infiltration capacity. In addition, in the absence of tree roots that provide soil anchorage, the likelihood of soil erosion is heightened. The erosion of soil has the potential to impede the flow of water in rivers, streams and drainage systems, thereby intensifying the occurrence of flooding. The presence of barren and fallow lands, small streams or rivers and urbanisation in the north of Idukki increases the chances of floods in

the region and the flood map suggests that the western region of Idukki, which constitutes high settlements, are high flood-prone areas.

4.8. Impact of distance from river

This parameter plays an important role in assessing flood-prone zones because Periyar river flows in Idukki District and is considered to have the largest discharge potential in Kerala. Because of very little distance from the river and dense drainage networks western and central regions of Idukki are under very high flood risk categories in the final flood susceptibility map (Figure 5).

4.9. Impact of soil

This parameter explains the reason for frequent floods in Idukki experienced in the past. As indicated in the soil map (Figure 4(d)), most of the study region is covered with clayey and loam soil, which has extremely low rates of permeability, due to which water is unable to infiltrate into the ground and hence the majority of the water flows as surface runoff, which further increases the chances of flooding.

4.10. Impact of geology

This parameter indirectly controls the water infiltration rates. Approximately 50% of the region is covered with charnockite rock, while other major rocks in the region are biotite gneiss and granite gneiss. Charnockite is a type of metamorphic rock which is an impermeable rock. Similarly, biotite gneiss and granite gneiss rocks belong to the family of metamorphic rocks and hence are treated as impermeable rocks. This is due to the impervious nature of these rocks which explains why the majority of the areas are high to very high flood-prone regions in Idukki District.

4.11. Impact of rainfall

Rainfall plays a major part in production of flood susceptibility maps. From the rainfall map, it is clearly depicted that Idukki and Thodupuzha received the highest annual rainfalls and hence both regions are very high flood-prone zones. Higher magnitude of rainfall results in more surface runoff and hence increases the chances of flooding in a region.

4.12. Impact of aspect

Factors such as aspect determine the path of floods in flood susceptibility mapping. The aspect map is created from the DEM and further classified into 10 classes. With this map, the direction of flow of water is determined, which further assists in the production of the flood susceptibility map accurately in the Idukki district.

5. Sensitivity analysis

Sensitivity analysis is performed to determine the key factors that have the greatest impact on the development of the flood susceptibility map. Sensitivity analysis is a method employed to ascertain the impact of changes or shifts in specific inputs or parameters of a given system on the resultant outputs or outcomes of that system. The evaluation of the strength and dependability of models is a prevalent practice in decision-making procedures. The goal of performing sensitivity analysis is to comprehend the interrelationships and interdependencies that exist between the input parameters and the resultant output of a given model or system. The outcomes of sensitivity analysis facilitate the identification of crucial variables or parameters that exert a significant impact. These data can be utilised to allocate resources effectively in data-gathering, concentrate on the most significant variables and enhance the decision-making procedures. Additionally, it has the potential to offer valuable perspectives on the hazards and unpredictability linked to the system under examination. The present investigation conducts a sensitivity analysis by utilising the final set of weights obtained from the AHP model. According to the sensitivity analysis as shown in Figure 7, parameters such as rainfall, elevation, distance from river, slope and TWI are the key contributing variables, while aspect, geology, LULC and soil are the least important variables among all the flood conditioning parameters.

6. Discussion

In the current study, a GIS-based AHP approach was employed to assess the flood-prone regions in Idukki District and the sensitive parameters that assist in

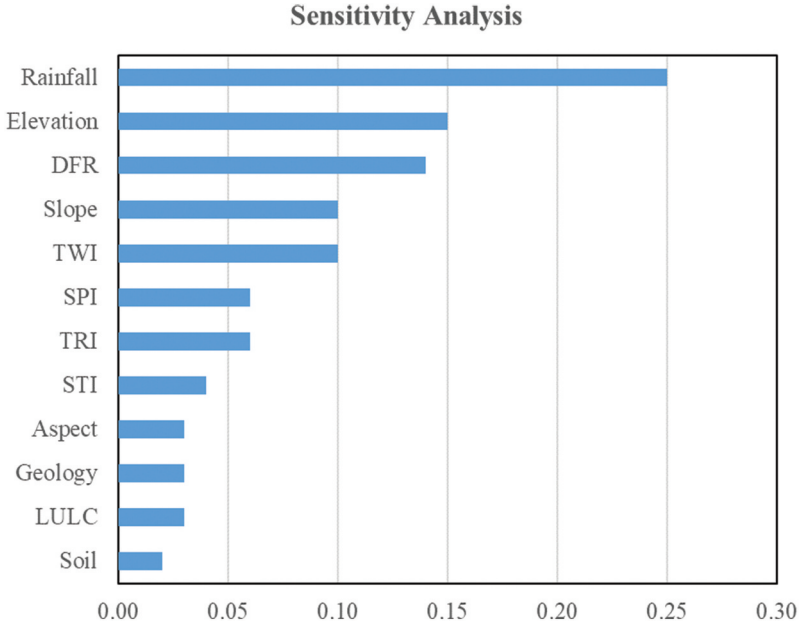


Figure 7. Variable importance of conditioning factors showing the sensitivity of each parameter.

development of an accurate flood susceptibility map. The efficacy and usage of the AHP model have been confirmed with several works in the domain of flood hazards and risk assessment (Liu *et al*. 2020, Pathan *et al*. 2022). Twelve flood conditioning parameters are utilised including topo-hydrological aspects such as TWI, TRI, STI and SPI in the delineation of the flood susceptibility map to perform detailed assessment on the different subdivisions or taluks of Idukki. Weighted parameters of different flood conditioning factors were used with the AHP model in the GIS platform in a linear way using geospatial technology. Sensitivity analysis shows the most influential parameters responsible for disastrous floods in the district. It is evident that rainfall is the primary factor contributing to flooding. The incidence of flooding can be attributed to the substantial rainfall, which is consistent with the sensitive parameter derived from the AHP model. Elevation is a highly significant parameter in the region, second only to rainfall. This is due to the presence of the Western Ghats, which influence the magnitude of floods by generating runoff from higher terrains. The distance from the stream is identified as a crucial factor for flood susceptibility, alongside the elevation. The areas adjacent to the streams are typically the first to experience flooding, and these regions are inundated and significantly impacted by the floods, as evidenced in the Idukki region. The areas in close proximity to bodies of water such as streams, rivers and reservoirs exhibited a higher degree of susceptibility to flooding.

The slope of a land surface plays a significant role in worsening the inundation process by enhancing the velocity of the water flow. Areas with steep slopes tend to produce floods with high velocity, while areas with gentle slopes are more susceptible to severe flooding, which can have a significant impact on the inhabitants of such terrain. Furthermore, the runoff originating from higher slopes can trigger landslides and transport substantial quantities of sediment to low-lying regions, exacerbating the socioeconomic conditions and normal livelihoods of affected communities. The development of TWI holds significant implications in the context of flooding, as it can impede the absorption of water into the ground in areas that have already been saturated with moisture due to precipitation, leading to surface runoff. This phenomenon will result in an amplification of the flood's magnitude. The SPI exhibits a direct correlation with flood events, as it serves as an indicator of the stream's potential energy where the flood velocity is expected to reach its peak during inundation. The TRI, STI and aspect parameters are utilised to assess the roughness of the terrain, transport of sediments and slope direction, respectively, which are indicative of the severity of flooding. The geology and LULC are subsequent factors in the analysis, as they provide insight into the permeability of a rock and the rate of infiltration that induce surface runoff and the land cover that may be significantly impacted during flooding, respectively, and the soil's porosity and permeability, which play a crucial role. The primary factor contributing to the occurrence of floods is excessive rainfall. However, the severity of the flooding is significantly amplified by the topographical characteristics of the area.

The importance of each sensitive parameter should be utilised by policy-makers as they contribute differently in preparation of a flood map. Das 2020 prepared the flood

susceptibility of whole of the Western Ghats using fewer parameters employing the AHP model, which indicates that the model is well suited for Indian states and districts.

Here are some key findings of the study:

- The AHP method was adopted to evaluate the flood-prone regions in the Idukki district of Kerala, India. It was observed that the AHP model was able to produce accurate results in development of a flood susceptibility map of Idukki. This also reveals the wide applicability of the AHP model in regions where heavy rainfall and floods occur every year (Khanday and Sujatha 2022).
- Twelve parameters are utilised to determine the flood-prone regions, which signifies that the model can handle a large number of parameters with accuracy and correctness.
- Results of the developed flood susceptibility map assisted in evaluating flood-prone regions in all the subdivisions of Idukki. Hence, accurate and detailed assessment of floods in the different taluks of Idukki District will help decision-makers to safeguard the livelihood of many inhabitants residing in these districts.
- The developed flood susceptibility map was validated using pre-flood and during-flood satellite images with the help of the AUC-ROC approach. This further consolidates the applicability of the AHP model in flood studies.

Though AHP is a flexible model for complex decision-making, sometimes calculations get cumbersome when the set of criteria is too large. Hence, the main limitation of the model involves the consistency check at each step to ensure the accuracy of the model. Another disadvantage associated with the model is that it does not allow one to add or remove a variable during the process of AHP.

7. Validation

The primary attribute of any proposed model is its ability to match with reality and produce significant outcomes. In order to validate the AHP flood susceptibility mapping, it is necessary to assess its success rate. The historic flood areas were compiled using information from the Survey of India, and pre- and post-flood Sentinel-1 images. For the validation procedure, one SAR image before the actual flood event (17 May 2018) and another SAR image during the flood event (21 August 2018) in VV polarisation were acquired. The Sentinel-1 images are subject to erroneous noise, which includes rigorous geometric and radiometric corrections, thermal effects and speckle. Prior to its application, any satellite data necessitate careful pre-processing. The pre-processing stage involves two key procedures: firstly, the correction of the orbital file to eliminate orbital noise; and secondly, the correction of thermal noise to eliminate noise in the data acquired by sensors during the satellite's data-collection process. The degradation of data quality in regions with low mean signal response detected by the Synthetic Aperture Radar (SAR) system, such as lakes, standing water and rivers, can be attributed to thermal noise. Radiometric calibration is employed to calibrate RADAR reflectivity (DN) to the backscattering coefficient (physical units) and is primarily utilised to compare SAR images obtained at different times. Terrain correction involves the conversion of Sentinel-1 SLC (Single Look Complex) data from slant range geometry to a map coordinate system and

the correction of distortions such as foreshortening, layover and shadowing effects. Following the pre-processing stage of Sentinel-1 VV SAR images, a speckle filter is implemented to achieve image smoothing for both before and during the flood event, as shown in Figure 8(a,b), respectively. To obtain the flooded region or points, both the SAR images are analysed using the image ratio technique with a threshold value of 1.2. The flooded SAR image is depicted in Figure 9, where white colour indicates flooded region and black indicates non-flooded areas. The obtained results are in coherence with the previous studies (Jacinth Jennifer *et al*. 2022, Parthasarathy *et al*. 2021) and are validated using the maps published by Kerala State Disaster Management Authority. This procedure assisted in acquiring actual historic flood locations of the Idukki district in a raster format. The data are then vectorised and transformed to a polygon, from which random points for both flooded locations and non-flooded locations are created using spatially balanced points in ArcGIS. The current study also aims to validate the performance and efficiency of the AHP model through the AUC-ROC technique (Figure 10). For validating the developed flood susceptibility map, the area under the curve - receiver operating characteristics (AUC-ROC) were plotted to validate the flood susceptibility mapping for the AHP model with that of the flooded points. In the AUC-ROC curve, the 'X' axis indicates the false positive rate and the 'Y' axis the true positive rate. Values of AUC-ROC vary from 0 to 1, where 0.50–0.60 indicates poor, 0.61–0.7 indicates average, 0.70–0.8 indicates good, 0.81–0.9 indicates very good and 0.91–1 indicates excellent sensitivity. AUC is useful as it assists in computing the performance of prediction accuracy

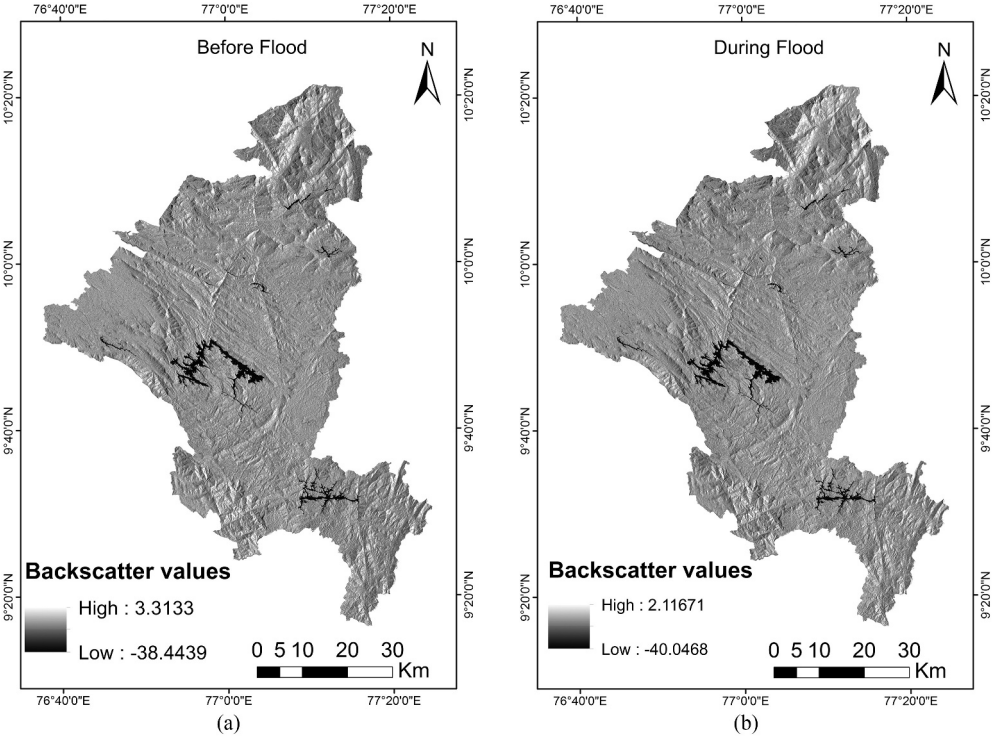


Figure 8. (a) SAR image before flood event. (b) SAR image during the flood event.

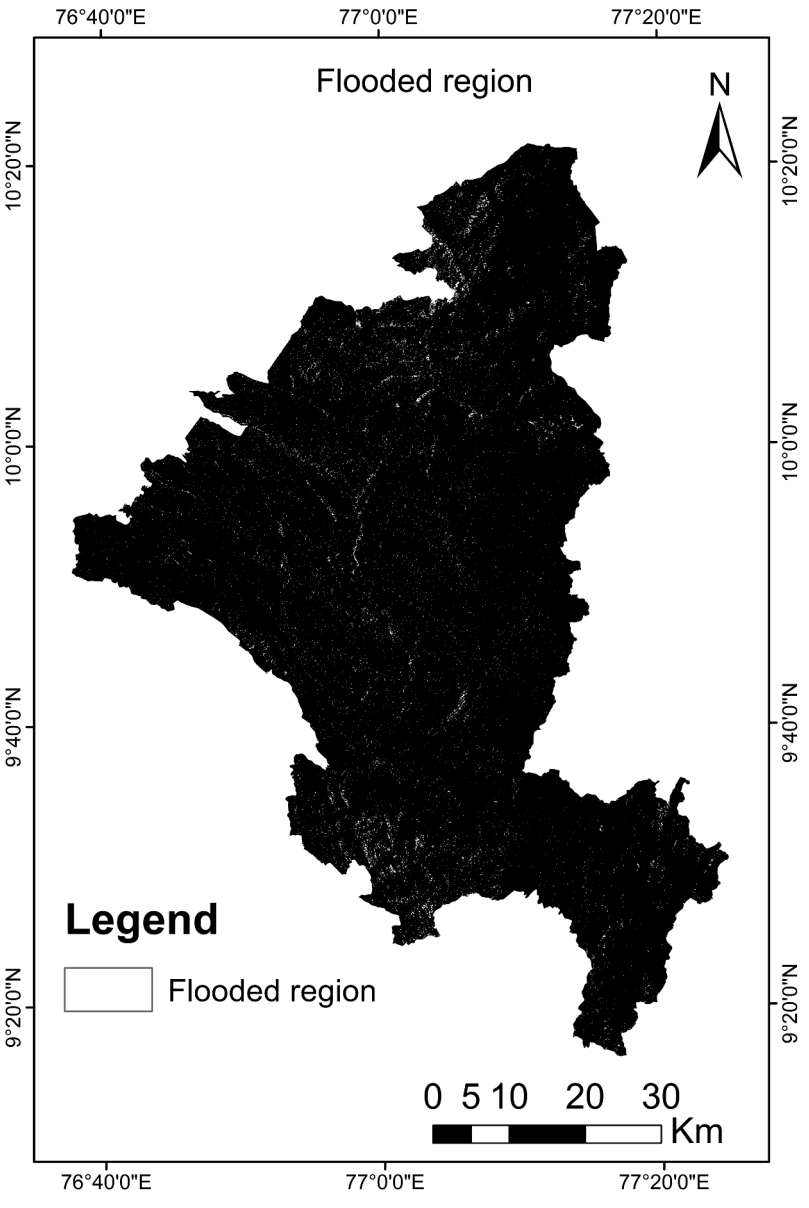


Figure 9. SAR image of flooded region of Idukki district.

and ROC is an effective method in assessing the performance of the model. The AUC value for the present study is evaluated as 0.91 or 91%, which is remarkably good accuracy for the employed spatial mapping model for flood susceptibility studies. This signifies that the adopted flood susceptibility approach can very well forecast the likelihood that a region will be flooded accurately. It seems that the proposed AHP approach can assist in developing flood susceptibility maps for other study areas also as this approach is not dependent upon the scale of the study region.

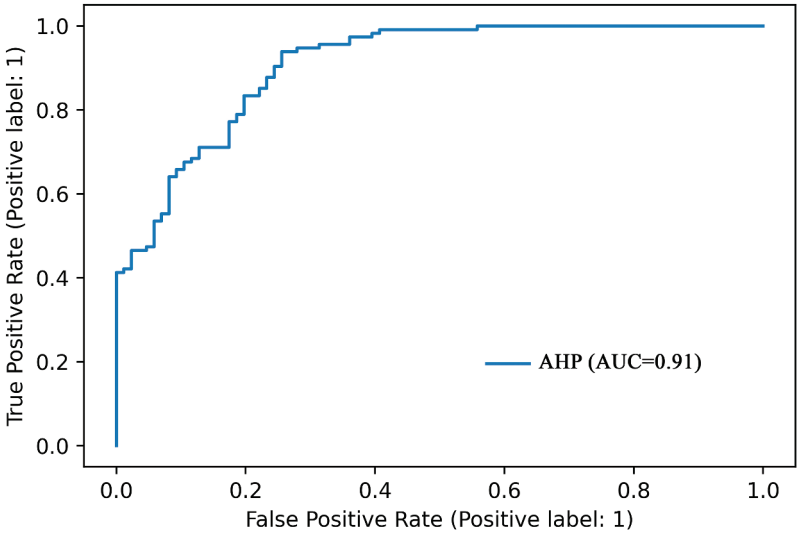


Figure 10. Validation of the flood susceptibility map using AUC-ROC curve.

8. Conclusion

The analytical hierarchy process (AHP) and geospatial techniques were utilised to demarcate flood-prone regions in the Idukki district of Kerala, resulting in a time-consuming and cost-effective methodology. Flood conditioning parameters such as slope, distance from the river, land use, topographic wetness index, topographic roughness index, rainfall, stream power index, sediment transport index, geology, elevation, aspect and soil map are all utilised in constructing a flood susceptibility map of the Idukki district in the ArcGIS platform. The developed flood susceptibility map reveals that about 1346.4 km^2 (~30% of the total area) was extremely susceptible to floods. These regions are mainly found in the western, central and southern sections of the Idukki district, where the slope grade rapidly changes and numerous minor streams join the main river. The accuracy of the AHP model was assessed with the help of the AUC-ROC approach, which indicated an accuracy of 91%. The approach utilised in this study has the potential to be used to provide flood-control recommendations in the study area.

Furthermore, the current study's future scope includes vulnerability and damage assessment of various subdivisions based on demographic criteria by assessing vulnerability index. The significance of each contributing parameter to floods and the optimum number of parameters to include in the AHP model should be tested against historic flood events of the region. For preparing susceptibility maps for flash floods, the AHP model has various benefits over other models, including fewer parameters, higher optimisation power and faster convergence. The utilisation of open-source data, like DEM, precipitation and soil characteristics, may present a limitation for the study due to their relatively low spatial resolution. The incorporation of high-resolution data, however, has the potential to enhance the understanding of flood dynamics and their corresponding variations in magnitude. Techniques for flood prevention have recently received more attention, particularly in Kerala, where flash

floods occur on an annual basis. Other basins and sites, on the other hand, have yet to undergo flood-control testing. As a result, the current study was conducted in the district of Idukki, employing AHP to give relevant information on ways that may be utilised to aid local governments and other parties in the implementation of successful flood mitigation measures and land-use policy planning.

Acknowledgements

The corresponding author is grateful to his supervisor Dr. Bharat Jhamnani, Assistant Professor, Delhi Technological University for his constant support and motivation during the research work. The author would also like to thank the laboratory staff of the Remote Sensing lab of the Delhi Technological University for their cooperation in carrying out the results for the current research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors reported there is no funding associated with the work featured in this article.

Data availability statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

References

Abbaszadeh, P., (2016). Improving hydrological process modeling using optimized threshold-based wavelet de-noising technique. *Water Resour Manage*, 30 (5), 1701–1721. doi:10.1007/s11269-016-1246-5.

Abdel Hamid, H.T., 2020. Environmental sensitivity of flash flood hazard using geospatial techniques. *Global Journal of Environmental Science and Management*, 6 (1), 31–46.

Abijith, D., *et al*., 2020. GIS-based multi-criteria analysis for identification of potential groundwater recharge zones - a case study from Ponnaniyaru watershed, Tamil Nadu, India. *HydroResearch*, 3, 1–14. doi:10.1016/j.hydres.2020.02.002.

Abijith, D. and Saravanan, S., 2021. Assessment of land use and land cover change detection and prediction using remote sensing and CA Markov in the northern coastal districts of Tamil Nadu, India. *Environmental Science and Pollution Research*, 29, 86055–86067.

Arabameri, A., *et al*., 2020. Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques. *Journal of Hydrology*, 587, 125007. doi:10.1016/j.jhydrol.2020.125007.

Aslam, M., *et al*., 2022. Adaptive machine learning based distributed denial-of-services attacks detection and mitigation system for SDN-enabled IoT. *Sensors*, 22 (7), 2697. doi:10.3390/s22072697.

Band, S.S., *et al*., 2020. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sensing*, 12 (21), 3568. doi:10.3390/rs12213568.

Benito, G., *et al*., 2010. The impact of late Holocene climatic variability and land use change on the flood hydrology of the Guadalentín River, southeast Spain. *Global and Planetary Change*, 70 (1–4), 53–63. doi:10.1016/j.gloplacha.2009.11.007.

Bradshaw, C., *et al*., 2007. Global evidence that deforestation amplifies flood risk and severity in the developing world. *Global Change Biology*, 13 (11), 2379–2395. doi:10.1111/j.1365-2486.2007.01446.x.

Bronstert, A., 2003. Floods and climate change: interactions and impacts. *Risk Analysis*, 23 (3), 545–557. doi:10.1111/1539-6924.00335.

Bui, D.T., *et al*., 2019. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *CATENA*, 179, 184–196. doi:10.1016/j.catena.2019.04.009

Chakraborty, S. and Mukhopadhyay, S., 2019. Assessing flood risk using analytical hierarchy process (AHP) and geographical information system (GIS): application in Coochbehar district of West Bengal, India. *Natural Hazards*, 99 (1), 247–274. doi:10.1007/s11069-019-03737-7.

Charlton, R., *et al*., 2006. Assessing the impact of climate change on water supply and flood hazard in Ireland using statistical downscaling and hydrological modelling techniques. *Climatic Change*, 74 (4), 475–491. doi:10.1007/s10584-006-0472-x.

Chen, J., *et al*., 2019. A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: a case study of the Yangtze River Delta, China. *International Journal of Environmental Research and Public Health*, 17 (1), 49. doi:10.3390/ijerph17010049.

Chen, Y.-R., Yeh, C.-H., and Yu, B., 2011. Integrated application of the analytic hierarchy process and the geographic information system for flood risk assessment and flood plain management in Taiwan. *Natural Hazards*, 59 (3), 1261–1276. doi:10.1007/s11069-011-9831-7.

Danandeh Mehr, A. and Kahya, E., 2017. A pareto-optimal moving average multigene genetic programming model for daily streamflow prediction. *Journal of Hydrology*, 549, 603–615. doi:10.1016/j.jhydrol.2017.04.045.

Danumah, J.H., *et al*., 2016. Flood risk assessment and mapping in Abidjan district using multi-criteria analysis (AHP) model and geoinformation techniques, (cote d'ivoire). *Geoenvironmental Disasters*, 3 (1), 10. doi:10.1186/s40677-016-0044-y.

Das, S., 2017. Signatures of morphotectonic activities in western upland Maharashtra and Konkan region.

Das, S., 2018. Geographic information system and AHP-based flood hazard zonation of Vaitarna basin, Maharashtra, India. *Arabian Journal of Geosciences*, 11 (19), 576. doi:10.1007/s12517-018-3933-4.

Das, S., 2019. Geospatial mapping of flood susceptibility and hydro-geomorphic response to the floods in Ulhas basin, India. *Remote Sensing Applications: Society and Environment*, 14, 60–74.

Das, S., 2020. Flood susceptibility mapping of the Western Ghat coastal belt using multi-source geospatial data and analytical hierarchy process (AHP). *Remote Sensing Applications: Society and Environment*, 20, 100379. doi:10.1016/j.rsase.2020.100379.

Das, S. and Gupta, A., 2021. Multi-criteria decision based geospatial mapping of flood susceptibility and temporal hydro-geomorphic changes in the Subarnarekha River, India. *Geoscience Frontiers*, 12 (5), 101206. doi:10.1016/j.gsf.2021.101206.

Dodangeh, E., *et al*., 2020. Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. *Science of the Total Environment*, 705, 135983. doi:10.1016/j.scitotenv.2019.135983.

Dottori, F., *et al*., 2018. Increased human and economic losses from river flooding with anthropogenic warming. *Nature Climate Change*. doi:10.1038/s41558-018-0257-z.

Dou, X., *et al*., 2017. Flood risk assessment and mapping based on a modified multi-parameter flood hazard index model in the Guanzhong Urban Area, China. *Stochastic Environmental Research and Risk Assessment*, 32 (4), 1131–1146. doi:10.1007/s00477-017-1429-5.

El Jazouli, A., Barakat, A., and Khellouk, R., 2019. GIS-multicriteria evaluation using AHP for landslide susceptibility mapping in Oum Er Rbia high basin (Morocco). *Geoenvironmental Disasters*, 6 (1), 3. doi:10.1186/s40677-019-0119-7.

Farhan Ul Moazzam, M., *et al*., 2018. Analysis of flood susceptibility and zonation for risk management using frequency ratio model in District Charsadda, Pakistan. *International Journal of Environment and Geoinformatics*, 5 (2), 140–153. doi:10.30897/ijegeo.407260.

Gabriels, K., Willems, P., and Van Orshoven, J., 2022. A comparative flood damage and risk impact assessment of land use changes. *Natural Hazards and Earth System Sciences*, 22 (2), 395–410. doi:10.5194/nhess-22-395-2022

George, S.L., *et al*., 2022. A multi-data geospatial approach for understanding flood risk in the coastal plains of Tamil Nadu, India. *Earth*, 3 (1), 383–400. doi:10.3390/earth3010023.

Ghosh, A. and Kar, S., 2018. Application of analytical hierarchy process (AHP) for flood risk assessment: a case study in Malda district of West Bengal, India. *Natural Hazards*, 94. doi:10.1007/s11069-018-3392-y.

Giovannettone, J., *et al*., 2020. Spatial analysis of flood susceptibility throughout Currituck County, North Carolina. *Journal of Hydrologic Engineering*, 25 (8), 05020021. doi:10.1061/(ASCE)HE.1943-5584.0001948.

GSI, 2006. Geological survey of India: district resource map of Kerala. https://www.gsi.gov.in/webcenter/portal/OCBIS

Gupta, L. and Dixit, J., 2022. A GIS-based flood risk mapping of Assam, India, using the MCDA-AHP approach at the regional and administrative level. *Geocarto International*, 37 (26), 11867–11899. doi:10.1080/10106049.2022.2060329.

Hadipour, V., Vafaie, F., and Deilami, K., 2020. Coastal flooding risk assessment using a GIS-based spatial multi-criteria decision analysis approach. *Water*, 12 (9), 2379. doi:10.3390/w12092379.

Hamlat, A., *et al*., 2021. Flood hazard areas assessment at a regional scale in M'zi wadi basin, Algeria. *Journal of African Earth Sciences*, 182, 104281. doi:10.1016/j.jafrearsci.2021.104281.

Hong, H., *et al*., 2018. Flood susceptibility assessment in Hengfeng area coupling adaptive neuro-fuzzy inference system with genetic algorithm and differential evolution. *Science of The Total Environment*, 621, 1124–1141. doi:10.1016/j.scitotenv.2017.10.114.

Houze, R.A., *et al*., 2017. Multiscale aspects of the storm producing the June 2013 Flooding in Uttarakhand, India. *Monthly Weather Review*, 145 (11), 4447–4466. doi:10.1175/MWR-D-17-0004.1.

Hsu, T.-W., *et al*., 2017. A study on coastal flooding and risk assessment under climate change in the Mid-western coast of Taiwan. *Water*, 9 (6), 390. doi:10.3390/w9060390.

Hunt, K.M.R. and Menon, A., 2020. The 2018 Kerala floods: a climate change perspective. *Climate Dynamics*, 54 (3–4), 2433–2446. doi:10.1007/s00382-020-05123-7.

Jacinth Jennifer, J. and Saravanan, S., 2022. Artificial neural network and sensitivity analysis in the landslide susceptibility mapping of Idukki district, India. *Geocarto International*, 37 (19), 5693–5715. doi:10.1080/10106049.2021.1923831.

Jacinth Jennifer, J., Saravanan, S., and Abijith, D., 2022. Integration of SAR and multi-spectral imagery in flood inundation mapping – a case study on Kerala floods 2018Integration of SAR and multi-spectral imagery in flood inundation mapping – a case study on Kerala floods 2018. *ISH Journal of Hydraulic Engineering*, 28 (sup1), 480–490. doi:10.1080/09715010.2020.1791265.

Kazakis, N., Kougias, I., and Patsialis, T., 2015. Assessment of flood hazard areas at a regional scale 563 using an index-based approach and analytical hierarchy process: application in Rhodope–Evros564 region, Greece. *Science of the Total Environment*, 538, 555–563. doi:10.1016/j.scitotenv.2015.08.055

Khanday, Z. A. and Sujatha, K., 2022. Flood hazard mapping and assessment using geospatial techniques and AHP model based hazard zonation -a case study of coastal Taluks of Cuddalore District, Tamil Nadu, India. *Xi'an Shiyou Daxue Xuebao (Ziran Kexue Ban)/Journal of Xi'an Shiyou University*, 18, 185–204.

Koks, E.E., *et al*., 2015. Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. *Environmental Science & Policy*, 47, 42–52. doi:10.1016/j.envsci.2014.10.013.

Kraus, C.N., *et al*., 2019. Unraveling flooding dynamics and nutrients' controls upon phytoplankton functional dynamics in amazonian floodplain lakes. *Water*, 11 (1), 154. doi:10.3390/w11010154.

Krishnaraj, A. and Honnasiddaiah, R., 2022. Remote sensing and machine learning based framework for the assessment of spatio-temporal water quality in the Middle Ganga Basin. *Environmental Science and Pollution Research*, 29 (43), 64939–64958. doi:10.1007/s11356-022-20386-9.

Kron, W., 2005. Flood risk = Hazard • values • vulnerability. *Water International*, 30 (1), 58–68. doi:10.1080/02508060508691837.

Kumar, S., 2013. *Rapidly assessing flood damage in Uttarakhand, India*. World Bank, 97332. http://documents.worldbank.org/curated/en/724891468188654600/Rapidly-assessing-flood-damage-in-Uttarakhand-India

Lei, X., *et al*., 2020. GIS-based machine learning algorithms for gully erosion susceptibility mapping in a semi-arid region of Iran. *Remote Sensing*, 12 (15), 2478. doi:10.3390/rs12152478.

Lim, J. and Lee, K., 2017. Investigating flood susceptible areas in inaccessible regions using remote sensing and geographic information system. *Environmental Monitoring and Assessment*, 189 (3), 96. doi:10.1007/s10661-017-5811-z.

Liu, Y., *et al*., 2020. Fine-scale coastal storm surge disaster vulnerability and risk assessment model: a case study of Laizhou Bay, China. *Remote Sensing*, 12 (8), 1301. doi:10.3390/rs12081301.

Lyubimova, T., *et al*., 2016. The risk of river pollution due to washout from contaminated floodplain water bodies during periods of high magnitude floods. *Journal of Hydrology*, 534, 579–589. doi:10.1016/j.jhydrol.2016.01.030.

Mahato, S., *et al*., 2021. Field based index of flood vulnerability (IFV): a new validation technique for flood susceptible models. *Geoscience Frontiers*, 12 (5), 101175. doi:10.1016/j.gsf.2021.101175.

Mahmoud, S.H. and Gan, T.Y., 2018. Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East. *Journal of Cleaner Production*, 196, 216–229. doi:10.1016/j.jclepro.2018.06.047.

Mandal, B. and Mandal, S., 2018. Analytical hierarchy process (AHP) based landslide susceptibility mapping of Lish River basin of Eastern Darjeeling Himalaya, India. *Advances in Space Research*, 62 (11), 3114–3132. doi:10.1016/j.asr.2018.08.008.

Meles, M.B., *et al*., 2020. Wetness index based on landscape position and topography (WILT): Modifying TWI to reflect landscape position. *Journal of Environmental Management*, 255, 109863. doi:10.1016/j.jenvman.2019.109863.

Moghaddam, D.D., Pourghasemi, H.R., and Rahmati, O., 2019. Assessment of the contribution of geo-environmental factors to flood inundation in a semi-arid region of SWIran: comparison of different advanced modeling approaches. *In*: *Natural Hazards GIS-based spatial modeling using data mining techniques*. Cham: Springer, 59–78.

NBSS, 1996. Soil map of Kerala. National Bureau of Soil Science and Land Use Planning. https://nbsslup.icar.gov.in/

Neumann, B., *et al*., 2015. Future coastal population growth and exposure to sea-level rise and coastal flooding - a global assessment. *PLoS One*, 10 (3), e0118571. (L. Kumar, ed.). doi:10.1371/journal.pone.0118571.

Nones, M., 2019. Dealing with sediment transport in flood risk management. *Acta Geophysica*, 67 (2), 677–685. doi:10.1007/s11600-019-00273-7.

Nourani, V., *et al*., 2014. A new hybrid algorithm for rainfall-runoff process modelling based on the wavelet transform and genetic fuzzy system. *Journal of Hydroinformatics*, 16 (5), 1004–1024. doi:10.2166/hydro.2014.035.

Pal, S. and Ziaul, S., 2017. Detection of land use and land cover change and land surface temperature in English Bazar urban centre. *Egyptian Journal of Remote Sensing and Space Sciences*, 20 (1), 125–145.

Paranunzio, R., *et al*., 2022. Assessing coastal flood risk in a changing climate for Dublin, Ireland. *Journal of Marine Science and Engineering*, 10 (11), 1715. doi:10.3390/jmse10111715.

Parthasarathy, K.S.S., *et al*., 2021. Chapter 17 - Assessing the impact of 2018 tropical rainfall and the consecutive flood-related damages for the state of Kerala, India. *In*: *Disaster resilience and sustainability a*daptation for sustainable development. Elsevier, 379–395.

Pathan, A.I., *et al*., 2022. AHP and TOPSIS based flood risk assessment- a case study of the Navsari City, Gujarat, India. *Environmental Monitoring and Assessment*, 194 (7), 509. doi:10.1007/s10661-022-10111-x.

Paul, G.C., Saha, S., and Hembram, T.K., 2019. Application of the GIS-based probabilistic models for mapping the flood susceptibility in Bansloi Sub-basin of Ganga-Bhagirathi River and their

comparison. *Remote Sensing in Earth Systems Sciences*, 2 (2–3), 120–146. doi:10.1007/s41976-019-00018-6.

Pourali, S.H., *et al*., 2016. Topography wetness index application in flood-risk-based land use planning. *Applied Spatial Analysis and Policy*, 9 (1), 39–54. doi:10.1007/s12061-014-9130-2.

Pradhan, B. and Youssef, A.M., 2011. A 100-year maximum flood susceptibility mapping using integrated hydrological and hydrodynamic models: Kelantan River Corridor, Malaysia. *Journal of Flood Risk Management*, 4 (3), 189–202. doi:10.1111/j.1753-318X.2011.01103.x.

Pramanick, N., *et al*., 2022. SAR based flood risk analysis: a case study Kerala flood 2018. *Advances in Space Research*, 69 (4), 1915–1929. doi:10.1016/j.asr.2021.07.003.

Rafiei-Sardooi, E., *et al*., 2021. Evaluating urban flood risk using hybrid method of TOPSIS and machine learning. *International Journal of Disaster Risk Reduction*, 66, 102614. doi:10.1016/j.ijdrr.2021.102614.

Rehman, A., *et al*., 2021. Formal modeling, proving, and model checking of a flood warning, monitoring, and rescue system-of-systems. *Scientific Programming*, 2021, 1–17. doi:10.1155/2021/6685978.

Saaty, R.W., 1987. The analytic hierarchy process—what it is and how it is used. *Mathematical Modelling*, 9(3–5), 161–176. doi:10.1016/0270-0255(87)90473-8.

Saaty, T.L., 1980. *The analytic hierarchy process: planning, priority setting, resources allocation*. New York: McGraw.

Saha, A., *et al*., 2021. Flood susceptibility assessment using novel ensemble of hyperpipes and support vector regression algorithms. *Water*, 13 (2), 241. doi:10.3390/w13020241.

Saha, A.K. and Agrawal, S., 2020. Mapping and assessment of flood risk in Prayagraj district, India: a GIS and remote sensing study. *Nanotechnology for Environmental Engineering*, 5 (2), 11. doi:10.1007/s41204-020-00073-1.

Sampson, C.C., *et al*., 2015. A high-resolution global flood hazard model. *Water Resources Research*, 51 (9), 7358–7381. doi:10.1002/2015WR016954.

Saravanan, S. and Abijith, D., 2022. Flood susceptibility mapping of Northeast coastal districts of Tamil Nadu India using multi-source geospatial data and machine learning techniques. *Geocarto International*, 37, 1–30.

Seejata, K., *et al.,* 2018. Assessment of flood hazard areas using analytical hierarchy process over the Lower Yom Basin, Sukhothai Province. *Procedia Engineering*, 212, 340–347. doi:10.1016/j.proeng.2018.01.044.

Sholihah, Q., *et al*., 2020. The analysis of the causes of flood disasters and their impacts in the perspective of environmental law. *IOP Conference Series: Earth and Environmental Science*, 437 (1), 012056. doi:10.1088/1755-1315/437/1/012056.

Siam, Z.S., *et al*., 2022. National-scale flood risk assessment using GIS and remote sensing-based hybridized deep neural network and fuzzy analytic hierarchy process models: a case of Bangladesh. *Geocarto International*, 37 (26), 12119–12148. doi:10.1080/10106049.2022.2063411.

Singh, S., 2019. Analytical study of Kerala floods. *International Journal for Research in Applied Science & Engineering Technology*, 7 (VI), 1851–1855.

Stevaux, J.C., *et al*., 2020. Changing fluvial styles and backwater flooding along the Upper Paraguay River plains in the Brazilian Pantanal wetland. *Geomorphology*, 350, 106906. doi:10.1016/j.geomorph.2019.106906.

Sturzenegger, M., *et al*., 2019, June. Semi-automated regional scale debris-flow and debris-flood susceptibility mapping based on digital elevation model metrics and Flow-R software. *In*: *Association of environmental and engineering geologists; special publication 28*. Colorado, USA: Colorado School of Mines. Arthur Lakes Library.

Surwase, T., *et al*., 2019. Novel technique for developing flood hazard map by using AHP: a study on part of Mahanadi River in Odisha. *SN Applied Sciences*, 1 (10), n.p. doi:10.1007/s42452-019-1233-6.

Suthirat, K., *et al*., 2020. AHP-GIS analysis for flood hazard assessment of the communities nearby the world heritage site on Ayutthaya Island, Thailand. *International Journal of Disaster Risk Reduction*. doi:10.1016/j.ijdrr.2020.101612.

Swain, K.C., Singha, C., and Nayak, L., 2020. Flood susceptibility mapping through the GIS-AHP technique using the cloud. *ISPRS International Journal of Geo-Information*, 9 (12), 720. doi:10.3390/ijgi9120720.

Talukdar, S., *et al*., 2020. Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. *Stochastic Environmental Research and Risk Assessment*, 34 (12), 2277–2300. doi:10.1007/s00477-020-01862-5.

Tehrany, M.S., *et al*., 2017. GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques. *Geomatics, Natural Hazards and Risk*, 8 (2), 1538–1561. doi:10.1080/19475705.2017.1362038.

Tehrany, M.S., Kumar, L., and Shabani, F., 2019. A novel GIS-based ensemble technique for flood susceptibility mapping using evidential belief function and support vector machine: brisbane, Australia. *PeerJ*, 7, e7653. doi:10.7717/peerj.7653.

Tehrany, M.S., Pradhan, B., and Jebur, M.N., 2013. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, 504, 69–79. doi:10.1016/j.jhydrol.2013.09.034.

Thomas, A., *et al*., 2021. Landslide susceptibility Zonation of Idukki district using GIS in the aftermath of 2018 Kerala floods and landslides: a comparison of AHP and frequency ratio methods. *Journal of Geovisualization and Spatial Analysis*, 5. doi:10.1007/s41651-021-00090-x.

Tran, T.A., *et al*., 2017. Application of multi-parametric AHP for flood hazard Zonation in coastal lowland area of central Vietnam. *International Journal of Geoinformatics*, 13 (1), 23–34.

Van Oldenborgh, G.J., *et al*., 2016. The heavy precipitation event of December 2015 in Chennai, India. *Bulletin of the American Meteorological Society*, 97 (12), S87–S91. doi:10.1175/BAMS-D-16-0129.1.

Veerappan, R. and Sayed, S.I., 2020. Urban flood susceptibility zonation mapping using evidential belief function, frequency ratio and fuzzy gamma operator models in GIS: a case study of greater Mumbai, Maharashtra, India. *Geocarto International*. doi:10.1080/10106049.2020.1730448.

Vimal, M., *et al*., 2018. The Kerala flood of 2018: combined impact of extreme rainfall and reservoir storage. *Hydrology and Earth System Sciences*. doi:10.5194/hess-2018-480.

Wallemacq, P. and Guha-Sapir, D., 2015. *The human cost of natural disasters 2015: a global perspective*. Centre for Research on the Epidemiology of Disasters CRED.

Wu, L., He, Y., and Ma, X., 2020. Can soil conservation practices reshape the relationship between sediment yield and slope gradient? *Ecological Engineering*, 142, 105630. doi:10.1016/j.ecoleng.2019.105630.

Xu, K., *et al*., 2021. The importance of digital elevation model selection in flood simulation and a proposed method to reduce DEM errors: a case study in Shanghai. *International Journal of Disaster Risk Science*, 12 (6), 890–902. doi:10.1007/s13753-021-00377-z.

Young, J.C., *et al*., 2022. Social sensing of flood impacts in India: a case study of Kerala 2018. *International Journal of Disaster Risk Reduction*, 74, 102908. doi:10.1016/j.ijdrr.2022.102908.

Zhang, G., *et al*., 2017. Flood effect on groundwater recharge on a typical silt loam soil. *Water*, 9 (7), 523. doi:10.3390/w9070523.

Biomass Conversion and Biorefinery
https://doi.org/10.1007/s13399-023-04742-4

ORIGINAL ARTICLE

Elucidating the molecular mechanism of ultrasound-enhanced lipase-catalyzed biodiesel synthesis: a computational study

Karan Kumar[1] · Pallavi Patro[2] · Unnati Raut[3] · Varsha Yadav[4] · Lepakshi Barbora[1] · Vijayanand S. Moholkar[1,5]

Received: 30 May 2023 / Revised: 7 August 2023 / Accepted: 9 August 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Sono-enzymatic biodiesel synthesis from different feedstocks has been significantly investigated in recent years, which reports the impact of ultrasound (US) on the structure of lipases. However, the links between alteration in the structure of lipases due to US and its activity enhancement have yet to be established. This study reveals the molecular mechanism of US-induced enhancement of lipase-catalyzed reactions in biodiesel synthesis. Using molecular simulations, we first analyzed the 3D structures and amino acid (aa) composition of the binding pockets of two lipases, viz. CALB and TLL. Next, docking analysis of different ligands (substrates and products of the reactions) with lipases was performed to reveal the interacting residues. We identified the US-induced alteration in the secondary structure of lipase from previous reports. Our analysis revealed that most binding pockets are located in the α-helix and random coil motifs, and binding interactions are mainly hydrogen bonding and hydrophobic interactions. We observed a rise in α-helix and random coils of lipase with sonication. The net effect of sonication is the widening of the catalytic cavity, thus easy substrate accessibility to the active site. Thus, enhanced enzymatic activity is manifested as faster kinetics of the reactions with higher yields.

Keywords Ultrasound · Lipase · Molecular docking analysis · Biodiesel · α-Helix · Random coil

1 Introduction

Biodiesel, derived from vegetable oils, animal fats, or waste cooking oils, is a renewable and sustainable alternative to conventional diesel fuel. It plays a crucial role in mitigating environmental concerns by reducing greenhouse gas emissions and emitting fewer pollutants [1, 2]. Biodiesel contributes to energy security and climate change mitigation by reducing dependence on fossil fuels and promoting a more sustainable energy future [3]. Recent research has focused on using microalgae as a sustainable feedstock for biodiesel production, addressing challenges in cultivation, harvesting, and extraction processes [3–5]. The combustion research of biodiesel and its model compounds has shown promising results in reducing emissions and improving combustion efficiency [6]. Techno-economic evaluations have highlighted the feasibility of biodiesel production using different catalyst technologies, with bulk calcium oxide catalysts being the most economically viable option [7]. These recent advancements in biodiesel research and technology provide valuable insights into the potential of biodiesel as a sustainable and environmentally friendly fuel source.

Conventionally, biodiesel is manufactured through an alkali-catalyzed transesterification process using vegetable oil or microalgal biomass as feedstock [2, 4, 8–10]. In the case of vegetable oils with a high free fatty acid content, acid-catalyzed esterification is also employed before transesterification. Enzyme-catalyzed biodiesel synthesis has also been explored in recent years [2, 11]. The most widely used enzyme for biodiesel synthesis is lipase.

Pallavi Patro, Unnati Raut, and Varsha Yadav contributed equally to this work.

✉ Vijayanand S. Moholkar
vmoholkar@iitg.ernet.in

[1] School of Energy Science and Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

[2] Department of Biochemistry, Pt. J.N.M Medical College, Raipur 492001, Chhattisgarh, India

[3] Department of Biological Sciences, Sunandan Divatia School of Science, Mumbai 400056, Maharashtra, India

[4] Department of Biotechnology, Delhi Technological University, Delhi 110042, India

[5] Department of Chemical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India

Published online: 21 August 2023

Conventionally, lipase, a versatile commercial enzyme, finds applications in pharmaceutical industries for bioremediation and environmental processes [12]. Lipase has also synthesized biodiesel from various feedstocks using esterification, transesterification, and interesterification [12–14]. Despite the merits of feedstock flexibility, high yields, and selectivity, the kinetics of reactions catalyzed by lipases is relatively slow, a significant bottleneck in their commercial applications in biodiesel synthesis [15]. The multifaceted applications of lipase make it a vital enzyme. Thus, its activity enhancement is crucial for industrial purposes. Numerous techniques have been attempted to enhance the kinetics of lipase-catalyzed reactions. One of these techniques is sonication or ultrasound (US) irradiation, which has proven effective for intensifying all routes to biodiesel synthesis, viz., esterification, transesterification, interesterification, and hydrolysis [16]. Sonication is a physical technique and does not involve toxic chemicals or harsh conditions and thus can be termed a "Green Technology" for enhancement of lipase-catalyzed reactions [15]. Recent studies [13, 17, 18] have revealed that the lipase enzyme activity increases significantly after exposure to the US with a concurrent reduction in the activation energy of the reactions.

The previous authors [13, 19] have reported the impact of US on the structure of lipases. However, the links between alteration in the structure of lipases due to sonication and its activity enhancement (which essentially is manifested in faster reaction kinetics) have yet to be established. Such investigation would require elucidation of the exact nature of the interaction of the substrates/products with active site residues of the lipase enzyme on a molecular level.

This study presents a comprehensive investigation into the molecular mechanism of sonication-induced enhancement in the kinetics of lipase-catalyzed reactions for biodiesel synthesis. By combining three approaches—determining the structure of binding pockets of lipases, visualization of binding pockets in different secondary structure motifs, and docking analysis of ligands with lipases using hybrid quantum mechanics/molecular mechanics (QM/MM) computations—novel insights are gained into the exact nature of interactions between substrates/products and lipase enzyme on a molecular level. This research contributes to a deeper understanding of sonication's impact on lipase activity, providing a general framework for analyzing ultrasonic enhancement of reactions catalyzed by other enzymes.

2 Methodology

The schematic representation of workflow and steps followed in the present study for elucidation of the molecular mechanism of the sonication-induced enhancement in the kinetics of lipase-catalyzed reactions is given in Fig. 1. Detailed procedure in a step-by-step manner followed in this study is presented in following sections.

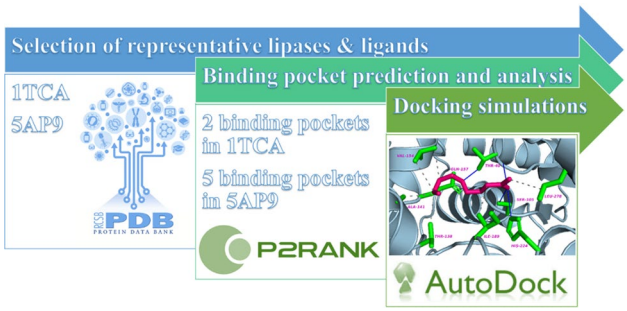


Fig. 1 Schematic representation of computational methodology followed in the present study

2.1 Selection of lipases and substrates for molecular docking experiment

For simulating esterification, transesterification, and hydrolysis reactions, we have chosen and imported two industrial lipases viz., *Candida antartica* lipase B (CALB; PDB ID: *1TCA*) and *Thermomyces lanuginous* lipase (TLL; PDB ID: *5AP9*), from the RCSB–PDB database [20]. These selected crystal structures were of good quality among others (judged based on the lowest atomic resolution and the Ramachandran scores provided in the PDB database). As per the requirement of ligands for molecular docking, we drew and optimized the chemical structure of the following compounds: methanol (MET), butanol (BOH), levulinic acid (LA), n-butyl levulinate (n-BL), aspirin methyl ester (AME), methyl salicylate (MSC), and oleic acid (OA) as representative substrates/products of esterification, transesterification, and hydrolysis reactions using density functional theory method in standalone Gaussian 16 software [21].

2.2 Sequence analysis and binding pocket prediction and molecular docking analysis

Before proceeding with molecular docking simulations, we first performed the sequence analysis and binding pockets prediction using 2StrctCompare (https://2struccompare.cryst.bbk.ac.uk/index.php) and PrankWeb (https://prankweb.cz/) software package of P2Rank web servers, respectively. The retrieved crystal structures (*1TCA* and *5AP9*) were visualized in PyMOL™ [22], and the secondary structure was compared using 2StrctCompare. The crystal structure of CALB and TLL was submitted to PrankWeb [23] to predict the

probable ligand-binding sites or binding pockets (BPs) in proteins. The high energies of CALB and TLL were eliminated using SWISS-PDB viewer (SPDBV) [24], and local minima closer to the native structure were achieved.

2.3 Molecular docking simulation and analysis

The molecular docking simulations of 7 ligands (described above) against both lipases were performed using Autodock v1.5.6 [22] linked with the MGLTools v1.5.6 viewer [23]. The ligand structure files were converted into "pdbqt" format using OpenBabel software. The procedure followed for molecular docking has been described in our earlier work [8]. The polar hydrogen ions and charges, such as Kollman and Gasteiger, were added to neutralize the system. A grid box was formed about coordinates for BP using the PrankWeb server. One hundred different docked conformations were generated using the Lamarckian Genetic Algorithm to investigate the optimal binding site of ligands to the lipases. These 100 docked conformations were ranked based on binding energies, and the best-docked conformation with the least binding energy was further analyzed in visualization software. The polar and non-polar interactions between protein–ligand complexes were studied in AutoDock, PyMOL™ version 2.4.1 [19] and PLIP webtool (https://plip-tool.biotec.tu-dresden.de/plip-web/). The 2D schematic representation of the enzyme-ligand complex's molecular interactions was analyzed using LigPlot + v.2.2 [24]. This protocol was reiterated for every ligand, and the final optimal conformation possessing the lowest binding energy among the hundred conformations procured for every operation was analyzed.

3 Results and discussion

3.1 Structural analysis and binding pocket characteristics

As outlined in the methodology section, the retrieved crystal structures of CALB and TLL were first visualized, and their sequences were compared using 2StructCompare. Both the lipases, CALB and TLL, are of α/β type fold and comprise 317 and 269 amino acid residues, respectively [25, 26]. DSSP (Database of Secondary Structure Assignments) algorithm for the prediction of secondary structures in CALB resulted in 37.5% helix (H), 12.3% extended sheets (E), and 50.2% random structure (R), which include turn and coils. On the contrary, TLL consists of 32% helix, 21.2% extended sheets, and 46.8% random structures.

The PrankWeb server identified the presence of two BP in CALB and five BP in TLL (refer to Table 1). The active site of CALB consists of SER_{105}-HIS_{224}-ASP_{187} with THR_{40}-GLN_{157} as an anionic hole contributing to the stabilization of the transition state. On the contrary, SER_{146}-HIS_{258}-ASP_{201} is present in active site in TLL. These active site residues are demonstrated in Fig. 2A(i, ii) in blue color. Figure 2A(i, ii) also demonstrate the residues in substrate/product BP in CALB and TLL, respectively. The amino acid composition of BP and results of 2D structure analysis using the DSSP algorithm are provided in Table 1. The minimap of secondary structure comparison produced using the 2StructCompare web server for CALB and TLL is provided in Fig. 2B.

Out of the multiple BPs in CALB and TLL, the pocket with the highest score was selected for molecular docking studies. We observed the presence of the glycine, leucine, valine, and

Table 1 Prediction of binding pockets for various commercial lipases using PrankWeb

PDB ID	No. of BP	Score	Composition of the binding pocket	2D structure analysis using the DSSP method		
				% H	% E	% R
1TCA	2	22.13	Chain A—THR_{40} TRP_{104} SER_{105} ASP_{134} THR_{138} LEU_{140} ALA_{141} LEU_{144} VAL_{149} VAL_{154} GLN_{157} GLU_{188} ILE_{189} HIS_{224} LEU_{278} ALA_{281} ALA_{282} ILE_{285}	37.5	12.3	50.2
		1.51	Chain A—ASP_{145} TRP_{155} THR_{158} GLY_{160} SER_{161} GLU_{294} ARG_{309}			
5AP9	5	6.29	Chain A—GLY_{61} ASP_{62} VAL_{63} THR_{64} ARG_{81} GLY_{82} SER_{83} ARG_{84} ASN_{92} LEU_{93} PHE_{113} SER_{116} LEU_{147} LEU_{151}	32.0	21.2	46.8
		6.19	Chain B—HIS_{110} GLY_{112} PHE_{113} SER_{116} LEU_{147} LEU_{151} ASP_{62} ARG_{81} GLY_{82} SER_{83} ARG_{84} SER_{85} TRP_{89} ASN_{92} ASN_{94}			
		1.47	Chain B—VAL_{120} THR_{123} LEU_{124} TYR_{53} PHE_{55} SER_{58}			
		1.22	Chain A—THR_{153} ALA_{173} PRO_{174} VAL_{176} ARG_{195} SER_{214} HIS_{215} SER_{216}			
		0.79	Chain B—ALA_{19} ALA_{20} CYS_{22} ILE_{34} PHE_{51} SER_{54} PHE_{55} GLY_{65} PHE_{66}			

BP binding pockets, *H* helix, *E* extended sheets, *R* random structures

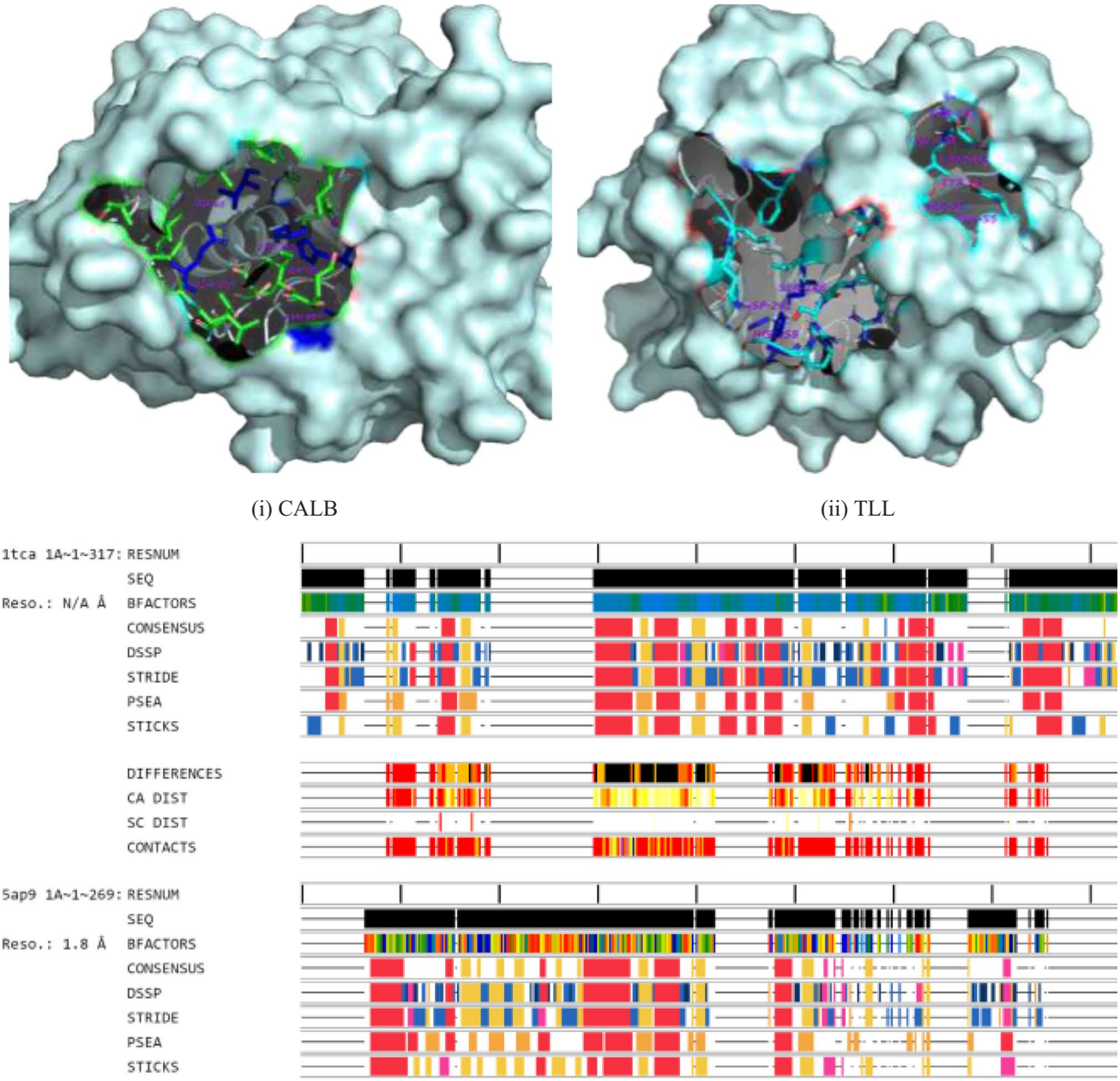


Fig. 2 **A** Visualization of most probable binding pockets of (i) CALB and (ii) TLL in PyMOL predicted by PrankWeb. Large cavities in the structure represent the highest-scored binding pockets of respective lipases listed in Table 1. Active site amino acid residues are represented in blue color. **B** Minimap of secondary structure comparison produced using 2StructCompare web server for both CALB and TLL

serine amino acid residues in most BPs (Table 1). The other amino acid residues in the BPs are alanine, glutamine, histidine, isoleucine, phenylalanine, and tryptophan (Table 1). The PyMOL visualization of the BPs (shown in Fig. 2A) revealed that most of these residues are in the random coil region of the enzyme. Figure 2A shows the position of the most probable BP (i.e., BP with high scores) for substrate and product during esterification/transesterification reactions. Previous studies have reported that α-helices are mainly involved in protecting and activating the lid region of the lipase [12, 13]. In contrast, β (or extended)-sheets are primarily engaged in the structural integrity of the enzyme structure [12, 16, 27]. The secondary structural comparison of CALB and TLL is shown in Fig. 2B.

3.2 Molecular docking analysis shows central involvement of α-helix and random coil in complex formation

Molecular docking studies were conducted to get insight into the molecular mechanism of lipase-catalyzed esterification,

transesterification, and hydrolysis reactions. Protocols and steps followed for docking simulations are provided in Section 2.2. The results summarized in Table 2 and Fig. 3 reveal the significant interactions in the lipase-ligand (here, ligand represents either substrate or product of esterification/transesterification/hydrolysis reaction) complexes. These interactions are of type hydrophobic bonds, hydrogen bonds, and salt bridge interactions. Most of the interacting amino acids were present in the α-helix and random coil domain of the lipase, as seen in Table 2. Docking system indices (DSIs) in Table 2 correspond to the type of ligands (i.e., substrates and products) with which the docking analyses of lipase enzymes have been performed, which maps with Table 3 for respective lipase-catalyzed reactions.

As discussed in the introduction section of this study, biodiesel synthesis occurs majorly through esterification or transesterification reactions. The role of CALB and TLL in reversible catalysis of esterification/transesterification/hydrolysis reactions has been extensively studied for biodiesel synthesis [2, 13, 19]. Table 3 summarizes the types of reactions studied in the present research. Tables 2 and 3 show that DSIs A1, A2, and A3 represent levulinic acid's esterification with n-butanol. Similarly, DSIs (B0, B6) and (A0, A6) represent the transesterification reactions of mixed non-edible oils with methanol catalyzed by TLL and rapeseed oil with methanol catalyzed by CALB, respectively. The DSIs (A4 and A5) represent the hydrolysis reaction of aspirin methyl ester catalyzed by CALB.

It can be seen from Table 2 that for esterification and transesterification reactions, alcohol (i.e., methanol and butanol) substrates have higher binding energy compared to organic acid (levulinic acid and oleic acid) substrates. This suggests that alcohols have a lower binding affinity towards lipases than organic acid. Therefore, during the ping-pong bi-bi mechanism, organic acid will bind first with lipase, followed by the binding of alcohol after the release of the first product. Additionally, ester products (i.e., n-butyl levulinate and FAME) have a low binding affinity towards lipases compared to their substrates (i.e., alcohols and organic acids). Hence, they will be faster from the [E-P] complex. Similar trends were also observed in the hydrolysis reactions (DSIs: A4, A5, B4, and B5), where the substrate's binding affinity was lower than the product's binding affinity.

Figure 3(i–xiv) depicts ligands' binding and types of interactions in the respective lipases. It is observed from Fig. 3 that in the BP, two substrates (acid and alcohols) bind in the two distinct domains (i.e., alcohol-binding domain and acyl-binding domain). These

Table 2 Molecular docking results of interaction of various ligands with CALB and TLL lipases viz., *1TCA* and *5AP9*

DSI	Definition	Binding energy (kJ/mol)	Amino acids involved in binding	Region of the protein	Type of interactions
A0	*1TCA* + MET	− 1.93	GLU_{188}, ALA_{225}	α-Helix and random coil	Hydrogen bonds
A1	*1TCA* + BOH	− 2.53	TRP_{104}, ILE_{189}, LEU_{278}, ALA_{281}, THR_{40}, SER_{105}, HIS_{224}	α-Helix, random coil, β-sheets	Hydrophobic, hydrogen bonds
A2	*1TCA* + LA	− 2.91	GLY_{288}, LYS_{290}	Random coil	Hydrogen bonds, salt bridge
A3	*1TCA* + n-BL	− 4.04	THR_{138}, ALA_{141}, VAL_{154}, GLN_{157}, ILE_{189}, LEU_{278}, THR_{40}, SER_{105}, GLN_{157}, HIS_{224}	α-Helix, random coil, and β-sheets	Hydrophobic, hydrogen bonds
A4	*1TCA* + AME	− 4.35	THR_{138}, GLN_{157}, ILE_{189}, VAL_{190}, THR_{40}, SER_{105}, HIS_{224}	α-Helix, random coil, and β-sheets	Hydrophobic, hydrogen bonds, salt bridge
A5	*1TCA* + MSC	− 3.37	ILE_{189}, LEU_{278}, ALA_{281}, ILE_{285}, THR_{40}, SER_{105}, HIS_{224}	α-Helix and random coil	Hydrophobic, hydrogen bonds, salt bridge
A6	*1TCA* + OA	− 4.41	THR_{138}, ALA_{141}, GLN_{157}, ILE_{189}, VAL_{190}, LEU_{278}, ALA_{281}, ILE_{285}, THR_{40}, SER_{105}, HIS_{224}	α-Helix, random coil, and β-sheets	Hydrophobic, hydrogen bonds, salt bridge
B0	*5AP9* + MET	− 2.25	SER_{85}, GLU_{87}, ASN_{88}	Random coil	Hydrogen bonds
B1	*5AP9* + BOH	− 2.96	ARG_{81}, THR_{64}	Random coil	Hydrophobic, Hydrogen bonds
B2	*5AP9* + LA	− 4.43	SER_{85}, CYS_{86}, GLU_{87}, ASN_{88}	Random coil	Hydrogen bonds
B3	*5AP9* + n − BL	− 3.96	VAL_{63}, LEU_{93}, PHE_{113}, LEU_{151}, THR_{64}, SER_{116}	α-Helix and random coil	Hydrophobic, hydrogen bonds
B4	*5AP9* + AME	− 4.55	LEU_{147}, $THR_{64,}$ ARG_{81}, SER_{116}	α-Helix and random coil	Hydrophobic, hydrogen bonds
B5	*5AP9* + MSC	− 4.29	VAL_{63}, LEU_{147}, LEU_{151}, THR_{64}, ARG_{81}	α-Helix and random coil	Hydrophobic, hydrogen bonds
B6	*5AP9* + OA	− 4.28	VAL_{63}, TRP_{89}, ASN_{92}, LEU_{93}, PHE_{113}, LEU_{147}, LEU_{151}, SER_{83}, SER_{85}, ASN_{92}	α-Helix and random coil	Hydrophobic, hydrogen bonds

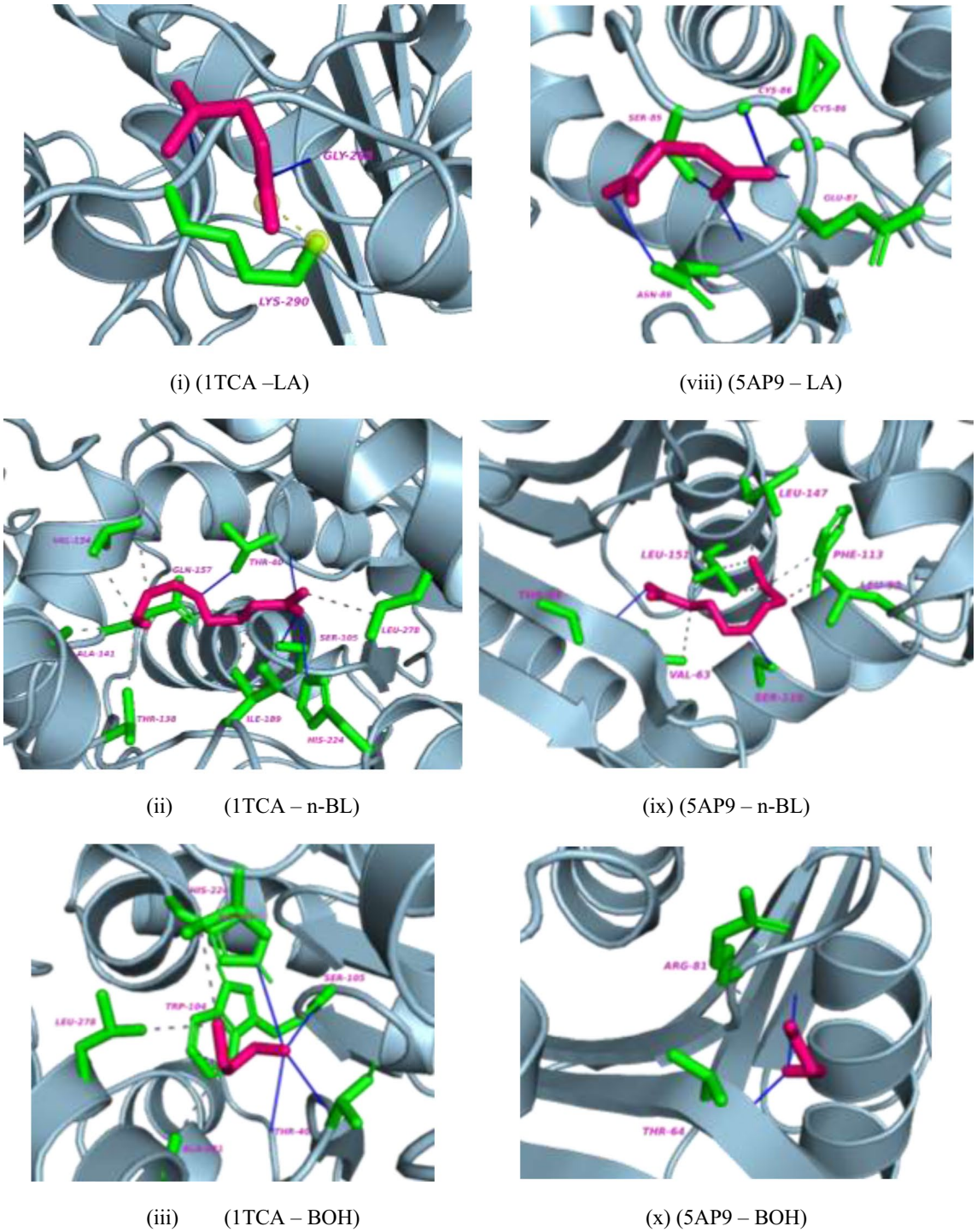
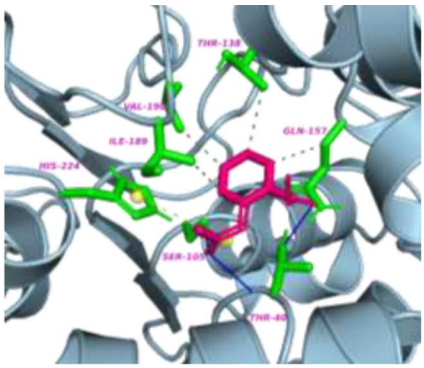
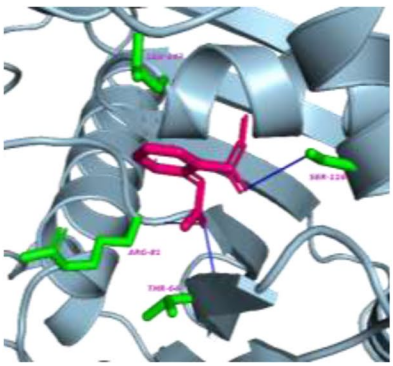


Fig. 3 Interaction profile of lipase CALB and TLL with respective substrates and products of esterification, transesterification, and hydrolysis reactions

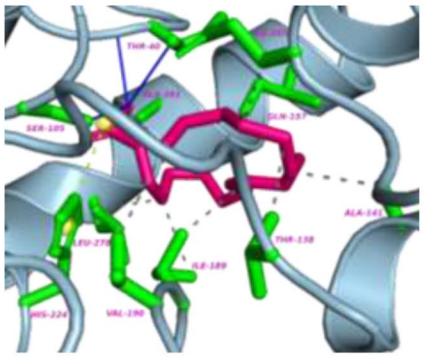
two binding domains help orient the substrate more effectively for catalysis to occur efficiently [13, 14]. Similar binding characteristics were observed in our study for esterification, transesterification, and hydrolysis reactions. It is evident from Fig. 3(i–xiv) that most of the enzyme-ligand binding and their



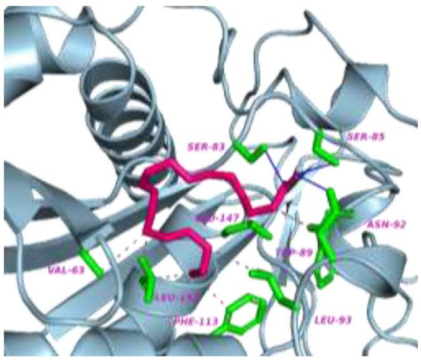
(iv) (1TCA – AME)



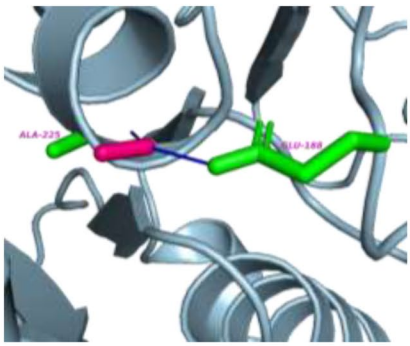
(xi) (5AP9 – AME)



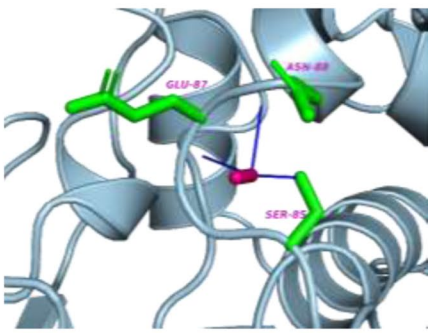
(v) (1TCA – OA)



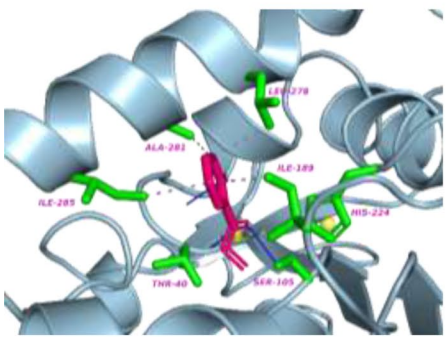
(xii) (5AP9 – OA)



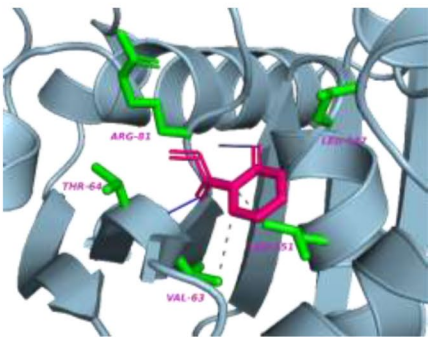
(vi) (1TCA – MET)



(xiii) (5AP9 – MET)



(vii) (1TCA – MSC)



(xiv) (5AP9 – MSC)

Fig. 3 (continued)

Table 3 Molecular docking analysis of representative lipase-catalyzed reactions for biodiesel synthesis

Experimental conditions				DSI*		Ref
Reaction type	Enzyme	Substrates	Product	ES complex	EP complex	
Esterification	Novozym 435	Levulinic acid and butanol	Butyl levulinate (BL)	A1, A2	A3	[13]
Transesterification	TLL	Non-edible oils and methanol	Fatty acid methyl ester (FAME)	B0, B6	–	[2]
Hydrolysis	CALB	Aspirin methyl ester (AME)	Methyl salicylate	A4	A5	[28]
Transesterification	Novozym 435	Rapeseed oil and methanol	Fatty acid methyl ester	A0, A6	–	[29]

*Refer to Table 2 for the docking system index (DSI) code

interaction happen in the random coil domain of the lipase.

3.3 Sonication-induced structural and morphological changes in lipase

3.3.1 Preamble

Ultrasound is a longitudinal wave that passes through any compressible medium through compression/rarefaction cycles. This induces small amplitude and high-frequency oscillatory motion of the fluid elements of the medium. Such oscillatory motion generates intense micro-mixing in the medium. The passage of ultrasound waves also induces variation in the bulk pressure in the medium. The secondary effect of sonication is cavitation–nucleation, growth, volume oscillations, and implosive collapse of gas or vapor bubbles driven by pressure variation induced by the ultrasound waves. The volume oscillations of cavitation bubbles also produce intense micro-convection near the bubbles. In addition, cavitation bubbles also give rise to acoustic waves due to the reflection of the fluid elements from the bubble/fluid interface. These physical phenomena also contribute to intense mixing in the medium. A peculiar feature of the convection generated by ultrasound and cavitation is its tiny spatial and temporal scales, i.e., this convection is caused at the micro-level (only a few micrometers) and for a very short duration (for example, the period of 50 µs for an ultrasound wave of 20 kHz). Such convection can induce micro-level phenomena in the medium, such as the unfolding of the protein molecules, which also leads to the rearrangement of the secondary structure of the protein molecules [30, 31]. A representative literature summary on the US-assisted enhancement of lipase-catalyzed trans-/inter-/esterification reactions for the synthesis of biofuels or their chemical additives is provided in Table 4.

3.3.2 Secondary structural analysis of lipase before and after sonication

In the context of the present study, numerous biochemical reactions catalyzed by lipases are accelerated due to sonication. A summary of representative recent literature on the enhancement effect of sonication (in terms of faster kinetics

Table 4 Summary of the representative literature on ultrasound enhanced lipase-catalyzed reactions

Reaction	Experimental conditions	Results	Ref
Esterification	CALB and *RML*; $T = 40$ ℃; $f = 20$ kHz; Sub.: FFA and xylitol; no solvent	Efficient synthesis of xylitol fatty ester (95% yield), with xylitol monoacyl ester and xylitol diacyl ester appearing as the main products (greater than 96%)	[32]
	Novozyme 435; $f = 35$ kHz; $T = 50$℃; $P = 35$ W; Sub.: LA and BOH; Sol: BME	Application of 35-kHz sonication boosted n-butyl levulinate yield from 70.9 to 92.22%	[13]
	Novozyme 435; $T = 50$℃; $P = 32$ mW; Sub.: AA and isomylOH	Application of US for 20 min, the ester concentration increased by 27.4% with higher enzyme reusability	[17]
	Novozyme 435; $f = 28$ kHz; $T = 50$℃; $P = 1.75$ kW/m^2; Sub.: butyric acid-lauric acid	Improved synthesis efficiency of butyric acid-lauric acid designer lipid by 11.38% with application of US	[6]
Transesteri-fication	TLL; $T = 36$℃; Sub.: non-edible oils and MeOH; $(MR)_{MET/oil} = 7.64$; EL = 3.55%	US + water addition to the reaction mixture (10% v/v) boosted biodiesel yield from 90 to 94%	[2]
	BCL; $f = 37$ kHz; $T = 60$℃; $P = 500$ W; Sub.: castor oil, EtOH, $(MR)_{MET/oil} = 6$; EL = 12 g, US probe	FAEE yield of 83% in 72 h with an ethanol-to-oil molar ratio of 3:1 in the case of the US	[33]
	Novozyme 435; $f = 28$ kHz; $T = 50$℃; Sub.: DHA and EPA	99% conversion to FAEE with sonication at the condition of 1 ml/min flow rate	[34]
Interesterification	CALB; P = 165W; $f = 25$ kHz; $T = 60$℃; Sub.: Crambe oil, MeOH; $(MR)_{MET/oil} = 12$	The US reduced total reaction time and percentage of enzyme loading by 20 wt%	[35]

Table 5 Influence of ultrasound on secondary structures of lipase: a literature summary

Reaction type	Experimental Conditions	Observed enhancement	Changes in secondary structures	Ref
Esterification	Novozyme 435; $f = 35$ kHz; T = 50°C; $P = 35$ W; duty cycle = 20%, time = 1 h; ultrasonic bath; Sub.: LA and BOH	Boosted conversion of LA from 70.9% to 92.22%	Increased α-helix and random coil; concurrent decrease in β-sheets	[13]
Hydrolysis	CVL; $f = 25.8$, 26.2, 35.5, 36.5 kHz; $T = 25$°C; $P = 29$–349 W, time = 20 min; ultrasonic bath; Sub.: Olive oil with PEG	The half-life of PEG400-treated lipase was 27-fold higher than untreated	No change in the circular dichroism spectra of the secondary structure	[41]
Enzyme activity	CALB and TLL; $f = 22$ kHz; $T = 50$°C; $P = 35$ W; duty cycle = 50–67%, time = 20–30 min; probe sonicator	The activity of CALB and TLL enhanced by 1.3- and 1.5-fold, respectively	Increase the random coil by 25.99 and 21.54% in both lipases; the rest components decreased	[42]
Hydrolysis	PPL; $f = 60$ kHz; $T = 37$°C; $P = 138$ W; duty cycle = 50–67%, time = 240 min; probe sonicator; Sub.: coconut oil	Lauric acid yield increased in the presence of sorbitol and mannitol osmolytes	Increased α-helix and random coil; concurrent decrease in β-sheets	[43]
Transesterification	BCL; $f = 40$ kHz; $T = 40$°C; $P = 120$ W; duty cycle = 50–67%, time = 2 h; ultrasonic bath; Sub.: IL and methanol	The best reaction media for biodiesel yield is 82.22%. Anionic IL was the most effective	Increased α-helix, β-sheets, β-turns; concurrent decrease in random coil and turns	[44]
Hydrolysis	Immobilized *alcalase*, $f = 22$, 28, 33, 40, 68 kHz; $T = 60$°C; $P = 600$ W each, time = 30–150 min; ultrasonic bath; Sub.: rapeseed protein	In US assisted-hydrolysis, hydrolysis increased by 74.38% at the optimum conditions	The number of a-helix and random coil decreased by 10.7% and 4.5%, β -chain increased by 2.4%	[45]

and higher yields) on different lipase-catalyzed reactions with changes in the 2D structural composition of lipase is presented in Table 5. It could be observed that these reactions involve other feedstocks and lipases from various sources. Moreover, operating conditions of sonication (such as frequency, intensity, duration) have also varied. Nonetheless, enhancement is seen in all cases, although the extent of enhancement varies from case to case. The ultrasound-induced enhancement in the kinetics of lipase-catalyzed reactions is a manifestation of the enzyme's enhanced activity, which is a consequence of the modifications in the secondary structure of the enzyme.

Previous authors have deduced the changes in the secondary structure of lipase due to sonication using different techniques such as circular dichroism, X-ray crystallography, SEM, deconvolution of FTIR spectra, NMR, and fluorescence spectroscopy [13, 27, 36–38]. The results of previous literature showing the alteration in the secondary structure of lipase are summarized in Table 5. As can be inferred from Table 5, the changes in secondary structure are highly lipase specific and a vital function of the operational conditions of sonication, such as frequency, intensity, and duration. However, on a broader basis, the significant changes in the secondary structure of lipase induced by ultrasound can be identified as (1) a rise in alpha helix content and (2) a rise in random coils. Few studies have reported a reduction in alpha helix content with a concurrent increase in beta sheets and turn. Our previous study [13] revealed that sonication increased the α-helix content that enhanced the activity of immobilized CALB by opening the flapping lid and widening the catalytic cavity hosting the binding pockets. Sonication at 35 kHz also caused a rise in the random coil content of lipase. In a similar study on the lipase-catalyzed synthesis of sucrose fatty acid ester, Zhang et al. [39] reported an increase in alpha helices in the enzyme structure of CALB using circular dichroism analysis. This shift in the secondary structure of CALB contributed to a 75% reduction in the reaction time.

As per the results presented in Tables 2 and 3 for two representative lipase enzymes CALB and TLL, the binding pockets of all ligands (viz. methanol, butanol, levulinic acid, n-butyl levulinate, aspirin methyl ester, methyl salicylate, and oleic acid) are mainly present in α-helix and random coils. Thus, a rise in α-helix and random coil content of lipase widens the catalytic cavity and relaxes the enzyme structure [8, 9], resulting in easy access of substrates to BP. Additionally, a rise in random coil content favors the formation of enzyme-ligand complexes (with a possible reduction in inhibition and unfavorable dissociation of intermediate complexes), ultimately enhancing the reaction kinetics and product yield. Similar observations were also found in our previous study [13]. On the other hand, increased beta sheets and turns contents caused reduced lipase activity, as

observed through CD analysis by Sun et al. [40]. In addition, SEM studies by Liu et al. [31] and Zhang et al. [39] revealed the presence of extra holes in the enzyme morphology, which suggest a rise in the surface area. This feature also contributes to faster enzymatic reaction rates.

The ternary approach of binding pocket analysis and concurrent analysis of the molecular docking simulations (Tables 2 and 3 and Fig. 3) in the present study and the changes in 2D structure of lipase due to sonication reported in the literature (Table 5) provide mechanistic insight into the beneficial influence of ultrasound on lipase-catalyzed reactions. The ternary approach developed in this study to deduce the molecular mechanism of US-induced enhancement in lipase-catalyzed reactions for biodiesel synthesis is schematically depicted in Fig. 4.

3.3.3 Suggestions for future work based on the present study

The present study attempts to offer a theoretical explanation for the ultrasound-induced enhancement of lipase-catalyzed reactions based on the ternary approach of determining the physical and chemical characteristics of the binding pockets of lipases, deducing their locations in different structural motifs and molecular docking analysis of other ligands using QM/MM simulations. Although this analysis could be plausible effort, an experimental confirmation of the computation results presented in this work requires to be made. The experimental study would require the lipase enzyme in free form (unlike the most commercially available enzymes in immobilized form). This work will also require high-end analytical instruments like X-ray crystallography, NMR solution, and Cryo-electron microscopy. The scientific community in enzyme research with access to these facilities can take up the future research in this area. We believe that our computational study will be a useful tool for designing such an experimental endeavor but also analysis of the results.

4 Conclusions

This study has attempted to provide mechanistic insight into the influence of sonication on lipase-catalyzed reaction with a ternary approach of determining the physical and chemical characteristics of the binding pockets of lipases, deducing their locations in different structural motifs and molecular docking analysis of other ligands using QM/MM simulations. It has been observed that most of the binding pockets are located in either α-helix or random coils. The previous literature has reported significant modifications in the secondary structure of lipases after exposure to sonication such as rise in α-helix and random coil contents. The net consequence of sonication is thus widening the catalytic cavity and easy accessibility for substrate binding. This is manifested in a rise in enzyme activity and faster kinetics of the lipase-catalyzed reactions. Therefore, the results of present

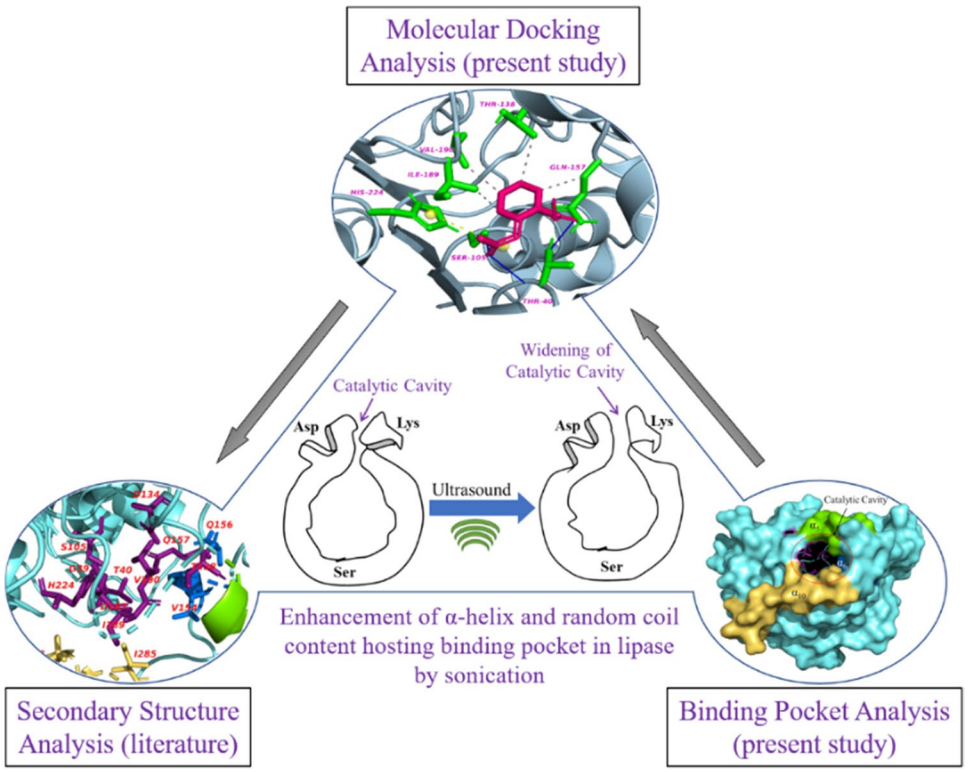


Fig. 4 Schematic depiction of the ternary approach to deduce the molecular mechanism of US-induced enhancement in lipase-catalyzed reactions for biodiesel synthesis

study have provided a mechanistic explanation to the previous literature on ultrasound-assisted intensification of esterification, transesterification, and hydrolysis reactions. Moreover, the methodology of ternary approach proposed in the present study also forms a general framework for analyzing the mechanism of ultrasonic enhancement of other enzyme-catalyzed reactions that are enhanced with sonication.

Abbreviations *1TCA*: PDB ID of *Candida antarctica* lipase B; *5AP9*: PDB ID of *Thermomyces lanuginosus*; MET: Methanol; BOH: Butanol; LA: Levulinic acid; n-BL: N-butyl levulinate; AME: Aspirin methyl ester; MSC: Methyl salicylate; OA: Oleic acid; CALB: Lipase B from *Candida antarctica*; TLL: Lipase from *Thermomyces lanuginosus*; Novozyme 435: CALB immobilized; RML: *Rhizomucor miehei* Lipase; Sub. : Substrate; FFA: Free fatty acid; BME: N-butyl methyl ether; US: Ultrasound; AA: Acetic acid; isomylOH: Asomyl alcohol; EL: Enzyme loading; *BCL*: *Burkholderia cepacia* Lipase; MR: Molar ratio; EtOH: Ethanol; DCM: Dicholoro methane; DHA: Docosahexaneoic acid; ES: Enzyme-substrate; EP: Enzyme-product; Sub. : Substrate; FFA: Free fatty acid; LA: Levulinic acid; IL: Ionic liquid; *PCL*: *Pseudomonas cepacia* Lipase; FTIR: Fourier transform infrared spectroscopy; QM/MM: Quantum mechanics and molecular mechanics; P: Power; T: Temperature; *f*: Frequency; CVL: *Chromobacterium viscosum* Lipase; PPL: Porcine pancreas lipase; CD: Circular dichroism; SEM: Scanning electron microscope; NMR: Nuclear magnetic resonance; FAEE: Fatty acid ethyl ester; FAME: Fatty acid methyl ester; EPA: Eicosapentaenoic acid

Acknowledgements Mr. Karan Kumar is grateful to the Prime Minister's Research Fellowship given by the Ministry of Education, Government of India. PP, UR, and VY acknowledge "Virtual INternship in Mathematical Biology: Modelling and SIMulations—2022 (VIMBIOSIM'22)" organized online at School of Energy Science and Engineering, Indian Institute of Technology Guwahati, Assam, India. The use of the HPC parallel computational facility, PARAM-ISHAN, at IIT Guwahati is highly acknowledged.

Author contribution K.K.: conceptualization, methodology, formal analysis, validation, software, visualization, investigation, writing—original draft, review and editing; P.P., U.R., V.Y.: methodology, formal analysis, investigation, software, writing—original draft; L.B.: supervision, writing—review; V.S.M.: supervision, project administration, writing—review and editing.

Data availability Not applicable.

Declarations

Ethical approval Not applicable.

Competing interests The authors declare no competing interests.

References

1. Malani RS, Pradhan S, Goyal A, Moholkar VS (2021) Mechanistic investigation in ultrasound-assisted interesterification using non-edible oil blends and heterogeneous catalyst. Asia-Pac J Chem Eng 16:e2638. https://doi.org/10.1002/apj.2638
2. Malani RS, Umriwad SB, Kumar K et al (2019) Ultrasound–assisted enzymatic biodiesel production using blended feedstock of non–edible oils: Kinetic analysis. Energy Convers Manag 188:142–150. https://doi.org/10.1016/j.enconman.2019.03.052
3. Ajala EO, Ajala MA, Akinpelu GS, Akubude VC (2022) Cultivation and processing of microalgae for its sustainability as a feedstock for biodiesel production. Niger J Technol Dev 18:322–343. https://doi.org/10.4314/njtd.v18i4.8
4. Singh N, Kumar K, Goyal A, Moholkar VS (2022) Ultrasound-assisted biodiesel synthesis by in–situ transesterification of microalgal biomass: optimization and kinetic analysis. Algal Res 61:102582. https://doi.org/10.1016/j.algal.2021.102582
5. Al-Sakkari EG, Attia NK, Habashy MM et al (2021) A bi-functional alginate-based composite for catalyzing one-pot methyl esters synthesis from waste cooking oil having high acidity. Fuel 306:121637. https://doi.org/10.1016/j.fuel.2021.121637
6. Liu W, Luo X, Tao Y et al (2022) Ultrasound enhanced butyric acid-lauric acid designer lipid synthesis: based on artificial neural network and changes in enzymatic structure. Ultrason Sonochem 88:106100. https://doi.org/10.1016/j.ultsonch.2022.106100
7. Gebremariam SN, Hvoslef-Eide T, Terfa MT, Marchetti JM (2019) Techno-economic performance of different technological based bio-refineries for biofuel Production. Energies 12:3916. https://doi.org/10.3390/en12203916
8. Maleki B, KalanakoppalVenkatesh Y, Siamak Ashraf Talesh S et al (2023) A novel biomass derived activated carbon mediated AC@ZnO/NiO bifunctional nanocatalyst to produce high-quality biodiesel from dairy industry waste oil: CI engine performance and emission. Chem Eng J 467:143399. https://doi.org/10.1016/j.cej.2023.143399
9. Maleki B, Esmaeili H (2023) Ultrasound-assisted conversion of waste frying oil into biodiesel using Al-doped ZnO nanocatalyst: Box-Behnken design-based optimization. Renew Energy 209:10–26. https://doi.org/10.1016/j.renene.2023.03.119
10. Maleki B, Esmaeili H, Mansouri M et al (2023) Enhanced conversion of dairy waste oil to biodiesel via novel and highly reactive UiO-66-NH2/ZnO/TiO2 nano-catalyst: Optimization, kinetic, thermodynamic and diesel engine studies. Fuel 339:126901. https://doi.org/10.1016/j.fuel.2022.126901
11. Dhawane SH, Kumar T, Halder G (2018) Recent advancement and prospective of heterogeneous carbonaceous catalysts in chemical and enzymatic transformation of biodiesel. Energy Convers Manag 167:176–202. https://doi.org/10.1016/j.enconman.2018.04.073
12. Filho DG, Silva AG, Guidini CZ (2019) Lipases: sources, immobilization methods, and industrial applications. Appl Microbiol Biotechnol 103:7399–7423. https://doi.org/10.1007/s00253-019-10027-6
13. Kumar K, Roy K, Moholkar VS (2021) Mechanistic investigations in sonoenzymatic synthesis of n-butyl levulinate. Process Biochem 111:147–158. https://doi.org/10.1016/j.procbio.2021.09.005
14. Kumar K, Moholkar VS (2023) Mechanistic aspects of enhanced kinetics in sonoenzymatic processes using three simultaneous approaches. In: Moholkar VS, Mohanty K, Goud VV (eds) Sustainable Energy Generation and Storage. Springer Nature Singapore, Singapore, pp 41–57. https://doi.org/10.1007/978-981-99-2088-4_5
15. Lerin LA, Loss RA, Remonatto D et al (2014) A review on lipase-catalyzed reactions in ultrasound-assisted systems. Bioprocess Biosyst Eng 37:2381–2394. https://doi.org/10.1007/s00449-014-1222-5
16. Javed S, Azeem F, Hussain S et al (2018) Bacterial lipases: a review on purification and characterization. Prog Biophys Mol Biol 132:23–34. https://doi.org/10.1016/j.pbiomolbio.2017.07.014
17. Vartolomei A, Calinescu I, Vinatoru M, Gavrila AI (2022) A parameter study of ultrasound assisted enzymatic esterification. Sci Rep 12:1421. https://doi.org/10.1038/s41598-022-05551-x

18. Wang D, Yan L, Ma X et al (2018) Ultrasound assisted interesterification of waste cooking oil and methyl acetate for biodiesel and triacetin production. Int J Biol Macromol 119:453–461. https://doi.org/10.1016/j.ijbiomac.2018.07.133
19. Bansode SR, Rathod VK (2017) An investigation of lipase catalysed sonochemical synthesis: a review. Ultrason Sonochem 38:503–529. https://doi.org/10.1016/j.ultsonch.2017.02.028
20. Zardecki C, Dutta S, Goodsell DS et al (2016) RCSB Protein Data Bank: a resource for chemical, biochemical, and structural explorations of large and small biomolecules. J Chem Educ 93:569–575. https://doi.org/10.1021/acs.jchemed.5b00404
21. Curtiss LA, Redfern PC, Raghavachari K (2007) Gaussian-4 theory. J Chem Phys 126:084108. https://doi.org/10.1063/1.2436888
22. Yuan S, Chan HCS, Hu Z (2017) Using PyMOL as a platform for computational drug design. WIREs Comput Mol Sci 7:. https://doi.org/10.1002/wcms.1298
23. Jendele L, Krivak R, Skoda P et al (2019) PrankWeb: a web server for ligand binding site prediction and visualization. Nucleic Acids Res 47:W345–W349. https://doi.org/10.1093/nar/gkz424
24. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723. https://doi.org/10.1002/elps.1150181505
25. Höck H, Engel S, Weingarten S et al (2018) Comparison of *Candida antarctica* lipase B variants for conversion of ε-caprolactone in aqueous medium—part 2. Polymers 10:524. https://doi.org/10.3390/polym10050524
26. Sandstrom AG (2010) Protein engineering of *Candida antarctica* lipase A: enhancing enzyme properties by evolutionary and semi-rational methods. Stockholm University, Department of Organic Chemistry
27. Khan FI, Lan D, Durrani R et al (2017) The lid domain in lipases: structural and functional determinant of enzymatic properties. Front Bioeng Biotechnol 5:. https://doi.org/10.3389/fbioe.2017.00016
28. Chiplunkar PP, Zhao X, Tomke PD et al (2018) Ultrasound-assisted lipase catalyzed hydrolysis of aspirin methyl ester. Ultrason Sonochem 40:587–593. https://doi.org/10.1016/j.ultsonch.2017.08.004
29. Firdaus MY, Brask J, Nielsen PM et al (2016) Kinetic model of biodiesel production catalyzed by free liquid lipase from *Thermomyces lanuginosus*. J Mol Catal B Enzym 133:55–64. https://doi.org/10.1016/j.molcatb.2016.07.011
30 Badgujar VC, Badgujar KC, Yeole PM, Bhanage BM (2021) Investigation of effect of ultrasound on immobilized *C. rugosa lipase*: synthesis of biomass based furfuryl derivative and green metrics evaluation study. Enzyme Microb Technol 144:109738. https://doi.org/10.1016/j.enzmictec.2020.109738
31. Liu S, Dong X, Wei F et al (2015) Ultrasonic pretreatment in lipase-catalyzed synthesis of structured lipids with high 1,3-dioleoyl-2-palmitoylglycerol content. Ultrason Sonochem 23:100–108. https://doi.org/10.1016/j.ultsonch.2014.10.015
32. Nieto S, Villa R, Donaire A, Lozano P (2021) Ultrasound-assisted enzymatic synthesis of xylitol fatty acid esters in solvent-free conditions. Ultrason Sonochem 75:105606. https://doi.org/10.1016/j.ultsonch.2021.105606
33. Moreira MAC, Alnoch RC, Luz Junior LF de L et al (2022) Transesterification of castor oil catalyzed by a fermented solid produced by *Burkholderia contaminans* using a bioreactor coupled with ultrasound irradiation. Biotechnol Appl Biochem n/a:https://doi.org/10.1002/bab.2410
34. Kuo C-H, Tsai M-L, Wang H-MD et al (2022) Continuous production of DHA and EPA ethyl esters via lipase-catalyzed transesterification in an ultrasonic packed-bed bioreactor. Catalysts 12:404. https://doi.org/10.3390/catal12040404
35. Tavares GR, Gonçalves JE, dos Santos WD, da Silva C (2017) Enzymatic interesterification of crambe oil assisted by ultrasound. Ind Crops Prod 97:218–223. https://doi.org/10.1016/j.indcrop.2016.12.022
36. Anobom CD, Pinheiro AS, De-Andrade RA et al (2014) From structure to catalysis: recent developments in the biotechnological applications of lipases. BioMed Res Int 2014:684506. https://doi.org/10.1155/2014/684506
37. Uppenberg J, Oehrner N, Norin M et al (1995) Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols. Biochemistry 34:16838–16851. https://doi.org/10.1021/bi00051a035
38. Zhang H, Sang J, Zhang Y et al (2019) Rational design of a *Yarrowia lipolytica* derived lipase for improved thermostability. Int J Biol Macromol 137:1190–1198. https://doi.org/10.1016/j.ijbiomac.2019.07.070
39. Zhang JC, Zhang C, Zhao L, Wang CT (2014) Lipase-catalyzed synthesis of sucrose fatty acid ester and the mechanism of ultrasonic promoting esterification reaction in non-aqueous media. Adv Mater Res 881–883:35–41. https://doi.org/10.4028/www.scientific.net/AMR.881-883.35
40. Sun T, Dong Z, Wang J et al (2020) Ultrasound-assisted interfacial immobilization of lipase on hollow mesoporous silica spheres in a Pickering emulsion system: a hyperactive and sustainable biocatalyst. ACS Sustain Chem Eng 8:17280–17290. https://doi.org/10.1021/acssuschemeng.0c06271
41. Talukder MMR, Shiong SCS (2015) Stabilization of *Chromobacterium viscosum* lipase (CVL) against ultrasound inactivation by the pretreatment with polyethylene glycol (PEG). Appl Biochem Biotechnol 177:1742–1752. https://doi.org/10.1007/s12010-015-1850-5
42. Nadar SS, Rathod VK (2017) Sonochemical effect on activity and conformation of commercial lipases. Appl Biochem Biotechnol 181:1435–1453. https://doi.org/10.1007/s12010-016-2294-2
43. Mojtabavi S, Hervé M, Forootanfar H et al (2022) A survey on the stabilizing effect of osmolytes on the ultrasound-irradiated lipase for efficient enzymatic hydrolysis of coconut oil. Colloids Surf B Biointerfaces 220:112910. https://doi.org/10.1016/j.colsurfb.2022.112910
44. Liu Y, Chen D, Yan Y et al (2011) Biodiesel synthesis and conformation of lipase from *Burkholderia cepacia* in room temperature ionic liquids and organic solvents. Bioresour Technol 102:10414–10418. https://doi.org/10.1016/j.biortech.2011.08.056
45. Wang B, Meng T, Ma H et al (2016) Mechanism study of dual-frequency ultrasound assisted enzymolysis on rapeseed protein by immobilized Alcalase. Ultrason Sonochem 32:307–313. https://doi.org/10.1016/j.ultsonch.2016.03.023

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-023-02051-7

ORIGINAL ARTICLE

Empirical validation of feature selection techniques for cross-project defect prediction

Ruchika Malhotra[1] · **Shweta Meena**[1]

Received: 2 April 2023 / Revised: 7 July 2023 / Accepted: 17 July 2023
© The Author(s) under exclusive licence to The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2023

Abstract In software engineering, cross-project defect prediction is an important area in the field of defect prediction and change prediction. Software defect prediction aims for the identification of defects in the early stages of the software development life cycle. Defect prediction helps in optimizing the resources in terms of testing and reduction of maintenance efforts. Defect prediction works appropriately if a huge amount of data is available for training a prediction model. Nowadays, a sufficient amount of data is not available due to which we have to use different data for training and testing of a project. The cross-project defect prediction idea emerged when different projects are used as training and testing dataset for the identification of defects. To design a defect prediction model we have to consider only significant features out of all the features set. Feature selection techniques are categorized based on unsupervised and supervised learning. The major limitation of cross-project defect prediction is handling different data distributions of source and target projects. The experiment was conducted using AEEEM and ReLink software defect dataset. Moreover, five projects of AEEEM and three projects of ReLink with a maximum count of files in the selected projects are 1862 and 399. In this study, we have analyzed the effect of feature selection techniques in cross-project defect prediction. The results were analyzed using AUC. The significance of filter, wrapper, and swarm search-based methods for feature selection techniques was analyzed separately. There is a trade-off between computational complexity and the performance of feature selection techniques. Swarm search-based methods performed better than filter and wrapper methods in terms of computational cost and overall performance of the prediction modes. The results were statistically validated using Friedman test and Wilcoxon signed rank test.

Keywords Defect prediction · Cross-project · Feature selection · Filter method · Wrapper method · Swarm search-based techniques

1 Introduction

Nowadays, prediction model plays an important role in Machine Learning (ML). In software engineering, software defect prediction is a very crucial domain. Defect prediction aims at the identification of defects at the initial stage of development, i.e., SDLC (Herbold 2015). Defect prediction plays an important role in improving the quality of software in terms of maintainability, functionality, reliability, and testability. To release defect-free software, every developer must check it from the criterion to maintain the quality of the software in upcoming years also. For designing a defect prediction model, it must be ensured that a sufficient amount of training data is used for designing prediction model. Moreover, the predictive model does not perform efficiently for future data. Prediction models are useful for predicting unseen data or future data. However, the unseen data should have similar characteristics to use a pre-designed prediction model.

In the existing studies, it has been demonstrated that the software defect prediction models are trained with a sufficient amount of training dataset and work efficiently for testing on the same dataset having similar distribution (Zou

✉ Shweta Meena
shwetameena@dtu.ac.in

Ruchika Malhotra
ruchikamalhotra2004@yahoo.com

[1] Department of Software Engineering, Delhi Technological University, Delhi 110042, India

Published online: 27 July 2023

et al. 2023). However, such type of prediction modes come under the category of Within-Project Defect Prediction (WPDP). In WPDP, the training and testing of the prediction model are accomplished using the same dataset. Moreover, the collection of a sufficient amount of training data is a serious issue.

The objective of conducting this study is to analyze the efficiency of FS methods for Cross-Project Defect Prediction (CPDP) using filter, wrapper, and search-based FS methods. In the existing, the authors have analyzed the individual performance of FS methods. However, according to the observation swarm based performed best for CPDP. Furthermore, the filter method selected the best features based on ranking criteria while swarm search methods are exploring optimal solutions. Moreover, swarm search-based methods explore the best feature subset out of all the subsets to find out the optimal solution.

In recent years, CPDP gained popularity among researchers, academicians, and developers. The idea of CPDP emerged from the availability of a sufficient amount of training data. In CPDP, two different projects are used for training and testing. One project is considered as a source project for training and another project is considered as a testing project for testing. The features of both projects must be checked. If the features of source and target projects are same then it is termed a Homogeneous Cross-Project (HoGCP). If the features of both projects are different, then it is termed a Heterogeneous Cross-Project (HeGCP) (Pal and Sillitti 2022).

In the case of HeGCP, both source and target projects may contain different attributes. To design prediction models, the attributes must be correlated. Furthermore, the matching metrics are selected based on the amount of correlation between different pairs of source and target dataset attributes. Moreover, transfer learning work when a prediction model is used for testing on some other project which is having similar characteristics as training data (Pan and Yang 2010). Moreover, the prediction model was learned through the knowledge provided by the source project. There are various types of transfer learning settings based on the domain of source and target dataset. However, transfer learning works perfectly if we want to use our pre-trained model for future projects. Three types of learning come under transfer learning such as feature transfer, instance transfer, parameter transfer, and relational knowledge transfer.

In the prediction model, before starting with model development, we have to preprocess our dataset. In the next phase, we extract relevant features of our dataset using various techniques termed Feature Selection (FS) techniques. FS techniques are categorized into three methods such as filter method, wrapper method, and swarm search-based method. FS helps in selecting the features which are relevant for model development concerning output variables (Bhutamapuram and Sadam 2022). The main aim of conducting this study is to analyze the impact of various FS techniques on CPDP. Every method has a different criterion, different threshold, different parameter settings, and aspects to select relevant features.

In today's era, the speed of accumulating data warehouses and databases is increasing rapidly, and dimensionality reduction is considered a major problem (Liu and Yu 2005). In further years, dimensionality reduction become a major problem for ML algorithms. Negative effects are also observed due to the scaling of the dimensionality of a dataset. The presence of irrelevant and redundant features adds noise to the dataset, which leads to a problem of overfitting (Hawkins 2004). Thus, data mining algorithms are developed in various areas such as text mining, bioinformatics, medicine, image processing, engineering, financial estimation, and sustainability. However, the significance depends on the ability to convert large amounts of data into acceptable form.

The cause of the defects in a software module is difficult to determine (Czibula et al. 2014). Various ML classification techniques are developed and used for defect prediction. The authors conducted studies to analyze the impact of FS and feature reduction techniques (Lessmann et al. 2008), and various cross-validation techniques in WPDP (Tantithamthavorn et al. 2017). In some of the existing studies, benchmark ML-based models were also proposed for WPDP such as logistic regression, Support Vector Machine (SVM), neural networks, k-nearest neighbor, decision trees, naïve bayes, etc.

In this study, the impact of each FS technique is analyzed statistically using filter, wrapper, and swarm search-based methods. The importance of the three methods of FS techniques is studied for improving software quality in terms of maintenance and effort estimation. AEEEM and ReLink datasets are considered in this study for CPDP. The projects used in this study are having similar metrics concerning AEEEM and ReLink. Five projects of AEEEM and three projects of ReLink are used in pairwise combinations of source and target pairs. The source and target pairs are created by combining different projects. Thus, this study was performed for 20 pairs of AEEEM datasets and 6 pairs of ReLink datasets. In this study, we have done a comparison between different FS methods along with the impact on CPDP. The experiment was conducted using WEKA 3.9.6. In existing studies, FS methods are not analyzed for CPDP. Filter methods work on the concept of ranker while wrapper and swarm search-based methods working is based on the principle of greediness.

This study aims to evaluate the effectiveness of different FS methods in conjunction with object-oriented metrics for CPDP. Furthermore, this study explored the importance of each FS method individually. Furthermore, the study also

explored different settings of source and target projects concerning the FS methods. Furthermore, ML classifiers are used for building prediction models. The results of all the predicted models are statistically validated for determining the best FS method for CPDP. The Research Questions (RQs) addressed in this study are:

RQ1: What is the predictive capability of filter methods for CPDP using AEEEM and ReLink dataset?
RQ2: What is the predictive capability of swarm search-based methods for CPDP?
RQ3: What is the predictive capability of filter, wrapper, and swarm search-based methods for CPDP?

1.1 Paper organization

Section 2 presents existing studies discussion on CPDP, FS techniques for software defect prediction, and transfer learning. The dataset used for experiments in this study is presented in Sect. 3. The research methodology used in this study is discussed in Sect. 4. Section 5 stated a detailed analysis of the results obtained in this study. Section 6 summarized the conclusions along with future work.

2 Related work

Prediction models designed for defect identification are learned from source datasets that consist of actual values and predictive metrics. In the existing studies, various supervised classifiers are used such as logistic regression (LR), decision trees, SVM, etc. to detect defect-prone parts of a software using training data.

In the existing study (Faiz et al. 2023), the authors conducted a study for developing a solution to select the optimal set of features for multi-class data using search-based optimization for CPDP. Moreover, the study considered quantitative and explanatory research types for experimentation. The classifiers used for optimization in this study are KN filter, ANN filter, Random Forest Ensemble (RFE), and Genetic Algorithm (GA). Further, *ckloc* is investigated as the best feature for improving the efficiency of a prediction model for multi-class problems. However, search-based optimization for FS provided 30–60% better results for multi-class CPDP. The software metrics represent the independent variables signified by product and process properties. In today's era, prediction models seem more interesting for academicians, researchers, and software experts (Rahman et al. 2012). The authors have used deep learning models for generating a large number of features and a significant amount of memory for reducing the computation time required for processing those features (Pramanik et al. 2023). A hybrid Genetic Algorithm (GA) and Grey Wolf Optimizer (GWO) are used to develop a prediction model by reducing a large number of feature sets.

In the existing study (Zhao et al. 2022), the authors experimented to resolve the problem of multi-source projects for defect prediction. Further, in the existing study, the author developed a Multi-Source-Based CPDP (MSCPDP) model using AEEEM and PROMISE dataset. The performance of the MSCPDP model is compared with the existing six CPDP methods showed F1 performance improvement by3.51%, 3.92%, 36.06%, 0.49%, 17.05%, and 9.49%, and the ACU median is improved by − 3.42%, 8.78%, 0.96%, − 2.21%, − 7.94%, and 5.13%.

To improve software quality, the volume of the training dataset and representativeness has a major effect on the usability and maintainability of the prediction model. The existing authors faced the quality issue in training data (Bird et al. 2009; Bachmann et al. 2010). However, some of the studies proved that the quality of the prediction model is as per benchmarks if it is properly trained on the noisy data as well (Kim et al. 2011).

In current years, CPDP gained popularity among software testing researchers. In the research world, the first study (Briand et al. 2002) was conducted in 2002 on CPDP, which discussed whether CPDP is worth studying or not. The researchers found that CPDP is a challenging problem and observed that the results are not significant. Furthermore, Zimmerman et al. (2009) conducted a study on defect prediction considering 622 pairs of cross-projects and the results were not significant. Out of 622 pairs, only 3.4% of cross-project pairs had significant performance in terms of precision, recall, and accuracy. Moreover, researchers and academicians have proposed various methods for improving CPDP.

Nam et al. (2013) proposed the TCA + method for CPDP. In CPDP, the common problem is the spatial distribution of features. Thus, TCA + defines some rules, and standardization strategy, and it performed transfer component analysis on CPDP to make similar distribution of source and target dataset features. In the existing study (Ma et al. 2012), a novel CPDP method Transfer Naïve Bayes (TNB) is proposed, which assigns weights for each instance in candidate source projects by predicting the distribution of target data. Some of the authors (Yu et al. 2017) analyzed the significance of features, parameters, and instances in CPDP. Empirical studies are conducted on various FS methods, proposing various feature matching and transfer of knowledge across CPDP (Yu et al. 2016). In existing studies, a compositional method of HYDRA is proposed for CPDP (Yu et al. 2019). In some of the existing studies (Wu et al. 2018), a Semi-Supervised CPDP (SSCPDP) problem was studied and a cost-sensitive kernel-based technique was proposed.

CPDP considers two different projects for source and target. Thus, the selection of appropriate variables is most

important. Relevant features can be selected by using various FS techniques. FS techniques are categorized into filter method, wrapper method, and embedded methods. The motivation behind using FS techniques is to compute the predictive capability of features or feature subsets to proceed with class labels. The filter method does not use any mining algorithm for FS. The simple idea of picking up the subset of features is used in the filter method (Visalakshi and Radha 2014). Filters methods are based on the general aspects of distance, information gain, dependency, and consistency (Liu et al. 2010). However, filter methods use ranking criteria for selecting relevant features. The filter method is independent of any ML classifier. Furthermore, the filter method preferred ranking criterion due to its simplicity, effectiveness, and good results. Furthermore, the advantage of using the filter method is to avoid overfitting, reduction in computational amount, fast speed, scalability for large datasets, and ease to understand.

Some of the commonly used filter methods are removing constant features, information gain, chi-square, and pearson correlation coefficient (Visalakshi and Radha 2014). FS is a NP-hard problem, to select the best set of features that provides the optimal solution. Secondly, the wrapper methods require data mining algorithms alike filter methods. The wrapper method selects relevant features to improve the performance of the mining algorithm. Clustering is used to select features using predictive capability (Bai et al. 2014). In the wrapper method, the different sets of features are searched. The quality of the selected attribute set is measured with a classification algorithm. Thus, the relationship is established between the feature set and prediction model in the wrapper method. In comparison to filter methods, wrapper methods are computationally more demanding and suffer from overfitting issues (Liu et al. 2010).

In the existing studies, researchers worked on FS techniques concerning WPDP. Furthermore, the researchers explored the performance of a combination of FS techniques and ensemble approaches for defect prediction (Laradji et al. 2015). However, the forward FS technique performed successfully well for the defect prediction model build for PC2, PC4, and MC1 imbalanced datasets. Moreover, some of the studies used the automatic FS technique for feature selecting appropriate features out of all the software metrics (Vashisht and Rizvi 2020).

To reduce computational cost and a number of features, the FS technique plays an important role in designing a prediction model. The performance of the model is also improved with appropriate FS techniques. In the case of the HeGCP prediction model, the set of software metrics is different, However, to select similar features from both the source and target dataset, FS techniques are used (He et al. 2013). However, the authors developed a model in combination with FS techniques using four ML classifiers followed by instance selection (Pal and Sillitti 2022).

Meta-heuristic methods used for FS are based on optimization. Embedded methods are described by flexibility, low computational cost, and simplicity. Embedded methods consider the number of candidate solutions encountered for the optimization of the feature set. Some commonly used embedded methods for FS are LASSO and RIDGE. The above methods that are searched based on strategies are categorized into simple, complete, and sequential models.

The FS techniques are explored for multiclass data also using the CPDP model. The authors (Faiz et al. 2023) used explanatory and quantitative approaches for experimentation. However, search based optimizer is used to obtain the best features set for CPDP. Thus, the authors observed 30–50% significant results in comparison with Genetic Instance Selection (GIS). The CPDP models are also developed using transfer learning by considering different projects for training and testing (Malhotra and Meena 2022).

In the existing study, the authors (Khatri and Kumar 2023) proposed an approach to deal with the limitations of CPDP through knowledge transfer. The proposed strategy of the authors is based on FS using iterative and non-iterative approaches. In the first strategy, the authors selected the important features and similar data distribution. The second strategy is the iterative strategy, which selects features by optimization based on performance. The authors (Tong et al. 2023) proposed a high-performance CPDP FS method. The novel FS method for CPDP is named Adaptive Triple Feature-Weighted Transfer Naïve Bayes (ARRAY). Moreover, it is concluded that ARRAY improved AUC by 6.5% and performed better than baseline methods for CPDP. ARRAY also performed better than baseline methods according to the posthoc test.

In the existing study authors (Vashisht and Murtaza 2023) proposed a CPDP model using a feature extraction approach. The feature extraction was done using a stacking-based ensemble learning method and auto-encoder. Further, two techniques are used to deal with an imbalanced dataset and computed the correlation among the features. However, the proposed CPDP framework provided better results by reducing 50% features and the highest AUC of 0.8901. In some of the research studies (Lyu et al. 2023), the researchers explored that filter-based techniques are popularly used in research.

Many of the research experts also stated that FS plays a significant role in the development of prediction models. FS is widely used when a huge number of features are available in the dataset used for experimentation. In CPDP, similar data distribution is a major problem in heterogeneous CPDP. Thus, the authors developed a CPDP model using optimal transport theory for establishing similarity in data distribution between source and

target projects (Zong et al. 2023). The developed model firstly measures the distance between two metrics spaces by learning an optimal transfer matrix. Semantic feature learning is also explored for software defect prediction (Liu et al. 2023). The authors used extracted natural language text in combination with programming language code for semantic feature learning. In the existing studies, the ML techniques developed from 2018 to 2021 are explored for software defect prediction (Sharma et al. 2023).

A novel hybrid algorithm Negative Binary General (NBG) (Feng et al. 2023), is proposed in combination with oversampling and FS algorithm for improving the performance of imbalanced classification. Researchers have also analyzed the performance of the Negative Positive Synthetic Minority Oversampling Technique (NPSMOTE) for improving sample generation practicability. However, Binary Ant Lion Optimizer (BALO) algorithm outperformed by extracting the most relevant features for the classification problem. The authors have used a novel stochastic approximation approach for performing simultaneous k-best Features Ranking (FR) and FS based on Simultaneous Perturbation Stochastic Approximation (SPSA) (Akman et al. 2023).

The authors developed a novel technique based on filtration methodology named as Relevance Diversity Algorithm (RDA) for FS (Shaheen et al. 2023). RDA is based on two measures relevance and diversity. RDA provides a low set of features by minimizing the amount of search time for relevant features. The authors (Tang et al. 2023) developed a novel FS technique via multiple graph fusion and feature weight learning. Thus, instead of learning the feature coefficient matrix, researchers can directly gain the weights of various feature dimensions through the feature weight matrix, and the weighted features are projected into the label space. The methodology considers evaluation of features subset for all the features, and finds out the optimal set of features according to th evaluation criteria (Li et al. 2017).

In some of the existing studies, wavelet transform is also used for FS. In the existing studies, SVM is compared with multilayer perceptron and fuzzy inference systems for FS. In their work, the total number of seventeen features are getting extracted by using the wavelet transform. By using LDA only four features are getting selected. The performance of the selected feature subset is compared with all features and features selected by Principle component analysis (PCA) which results in the selected feature subset giving better performance among these. In this work, SVM is also compared with the Multilayer Perceptron (MLP) and the Fuzzy Inference System (FIS) which results in the SVM classifier giving better performance than other classifiers.

3 Dataset used

The dataset characteristics, variables of the dataset, and dataset sources are discussed in this section. Furthermore, the category of the dataset, count, types of metrics, and number of buggy files in each dataset are summarized.

3.1 Dependent and independent variable

This study aims to analyze the effectiveness of FS techniques for CPDP. Thus, the efficiency of selecting features in the case of CPDP using filter, wrapper, and swarm search-based (embedded) methods is analyzed. The dataset used consists of independent and dependent variables. The dependent variables indicate the presence and absence of defects in a specific module such as buggy or clean. If the output of the dependent variable is clean, it indicates that the module is free from defects. Moreover, if the output of the dependent variable is buggy, it indicates the module consists of defects. In this study, two projects are used such as AEEEM and ReLink. The AEEEM dataset consists of 61 attributes as independent variables and 1 attribute as dependent variable. ReLink dataset consists of 26 independent variables and 1 dependent variable. The independent variables of the AEEEM dataset are grouped into five categories such as Source Code-Based Metrics (SCM) (17), previous defect-based metrics (5), entropy-based metrics (5), entropy of SCM (17), and churn of SCM (17). Table 1 discussed the five categories of the AEEEM dataset (Tong 2019; Hu and Zhu 2023). In Table 1 the first column depicts the dataset type, the second column depicts the metrics name, the third column discussed the metrics, and the fourth column indicates the number of metrics combined into five categories.

3.2 Dataset description

The experiment has been conducted using AEEEM and ReLink datasets in this study. The project metrics of the dataset are summarized in Tables 2 and 3. Furthermore, a detailed description of the dataset is discussed (You et al. 2016). In Tables 2 and 3 first column represents s. no. corresponding to each project, the second column represents the project name (abbreviated name), the third column represents the type of project, the fourth column represents the total number of files in each project, the fifth column represents the percentage of buggy instances, and the sixth column represents no. of metric in each project. AEEEM and ReLink datasets are selected for experimentation in this study, to perform validation of FS techniques on different projects, validity of dataset, and most important varied feature sets.

Table 1 Statistic of AEEEM defect dataset

Category	Type	Description	# of metrics grouped into categories
Category 1	ck_oo_cbo ck_oo_numberOfLinesOfCode	Coupling based metrics Coupling-based metrics consider line of code	17
Category 2	numberOfBugsFoundUntil numberOfCriticalBugsFoundUntil	Defect based metrics Defect-based metrics consider critical security level	05
Category 3	CvsEntropy CvsLogEntropy	Change-based entropy metrics Change-based logarithmic entropy metrics	05
Category 4	LDHH_cbo LDHH_numberOfLinesOfCode	Linearly decayed entropy of CBO metrics Linearly decayed entropy of LOC metric	17
Category 5	WCHU_cbo WCHU_numberOfLinesOfCode	Coupling-based metrics consider weighted churn of CBO metric Consider churn of LOC metric	17

Table 2 Dataset statistics of the data collected from AEEEM

S. no.	Project	Type	Number of files	% of buggy files	# of metrics
1.	Equinox (P1_EQ)	OSGi framework	325	36.69	71
2.	Eclipse JDT Core (P2_JDT)	IDE development	997	20.66	71
3.	Apache Lucene (P3_LUC)	Search engine library	39,691	9.26	71
4.	Mylyn (P4_MYL)	Task management	1862	13.16	71
5.	Eclipse PDE UI (P5_PDE)	IDE development	1492	14.01	71

Table 3 Dataset statistics of the data collected from ReLink

S. no.	Project	Type	Number of files	% of buggy files	# of metrics
1.	Apace (P6_AP)	Web server	194	50.52	26
2.	Safe (P7_SA)	Security	56	39.29	26
3.	ZXing (P8_ZX)	Bar-code reader library	399	29.57	26

4 Research methodology

The steps followed to experiment with this study are discussed in this section. However, the dataset considered for experimentation is summarized in Sect. 3. The research methodology is further represented graphically in Fig. 1.

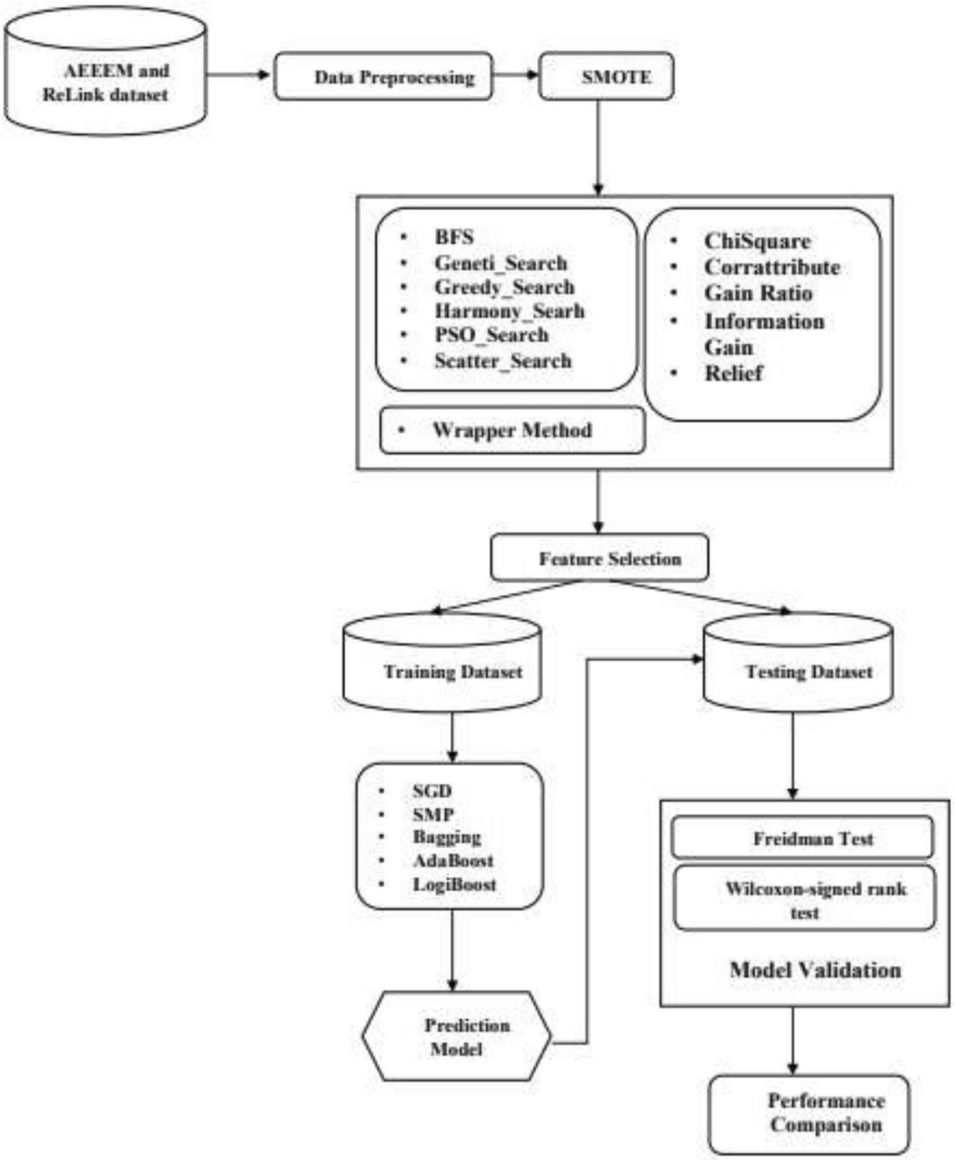
4.1 Data preprocessing techniques

In this study, the dataset is preprocessed before using for model development. For data preprocessing, missing values are handled, outliers are removed, null values are handled, and normalization of dataset techniques is performed. Firstly, the missing values or null values are replaced with the mean of the remaining values in the same column. Thus, the imputation of missing values by mean prevents information loss. Secondly, the outliers are handled using imputations. Moreover, the deletion of records due to outliers is not an advantage. In most cases, we try to retain the records by using some methods to prevent the loss of crucial information. However, the dataset needs to be normalized. However, normalization is a technique of converting data to a specified range of data. Furthermore, normalization techniques help in designing prediction models, statistical models, forecasting models, clustering models, and classification models. Thus, normalization is performed by converting all the numerical values in [0,1] considering the data used for normalization.

4.2 Imbalanced data

In a classification problem, when instances of one target class is more than the other class instances, this problem is termed an imbalanced dataset problem. There are various techniques exist to deal with imbalanced datasets such as SMOTE, oversampling, undersampling, and cost-sensitive learning. SMOTE is used in this study for balancing dataset. SMOTE is a technique, that oversamples the instances of minority class. Furthermore, new instances of minority class are synthesized from existing data only.

Fig. 1 Proposed methodology



4.3 Filter method

ChiSquare (FS1): It selects the relevant features based on ranking criteria. Furthermore, the chi-square test considers the independence of two events. Thus, the higher the value of chi-square, the attribute is depending on the outcome variable. Hence, the attribute will be selected by the chi-square test.

Correlation Attribute Evaluation (FS2): It selects the relevant attributes or features of a dataset using a correlation matrix. The correlation matrix predicts the amount of correlation between independent variables using pearson's correlation coefficient.

Gain Ratio (FS3): Gini ratio selects the relevant attributes by using a ranker. It assigns ranks to the features of high-dimension. Furthermore, the features with low ranks are removed and create reduces set of features. However, the K-anonymization privacy preservation techniques are used for both the original set of features and the reduced set of attributes.

Information Gain (FS4): Information gain is also used ranker for FS. Information gain provides the amount of useful information provided by each attribute for the outcome variable. It computes the gain of the attributes for the target variable. However, the feature with high gain is selected**.**

Relief Attribute Selection (FS5): The relief method is also based on the ranker criterion. It selects the relevant features according to the differentiation between instances that are near to each other. However, the relief method was designed for binary classification problems including non-metric data that is discrete or numerical features.

4.4 Wrapper method (FS6)

The selection of relevant features using the wrapper method follows a specific classifier that is used to develop a model. Unlike the filter method, wrapper method doesn't use a ranker. The wrapper method is based on the principle of greediness. It evaluates all the possible combinations for the selection of relevant attributes against the evaluation criterion. The evaluation measures for the regression problem are p-value and R-squared. The evaluation measures for classification problems are accuracy, precision, and recall. The commonly used wrapper methods are forward selection, backward selection, and bi-directional elimination.

4.5 Swarm search methods

In this study, six different swarm search methods are used for FS.

Best First Search (BFS) (FS7): BFS selects the attributes from an attribute set using search space. It selects the attributes by using greedy hill climbing augmented with a backtracking facility. Best first starts with an empty set of attributes and start searching for attributes in a specific direction by considering all possible combinations with additions and deletions.

Genetic Search (GENS) (FS8): GENS can be used for FS and hyperparameter optimization. It finds out the optimal set of attributes based on evolution.

Greedy Stepwise Search (GRS) (FS9): GRS selects the attributes using forward and backward search through the space of attributes subset. Starting criterion is either zero or all attributes or an arbitrary point in the feature space.

Harmony Search (HS) (FS10): It is inspired by the harmony improvisation process. HS method is considered a global optimization algorithm. In HS, at every iteration harmonies are generated termed as solution. Furthermore, these solutions are stored in harmony memory. HS method uses five parameters, three mandatory parameters, and two optional parameters.

Particle Swarm Optimization (PSO) (FS11): PSO method is an optimization approach based on the behavioral study of animals/birds. Furthermore, PSO performed better for FS for other techniques, Moreover, PSO has the capability of searching through large space, is computationally less expensive, and requires few parameters. The traditional PSO approach has some limitations. However, a new variant of PSO is developed. PSO (4–2) and BPSO overcome the limitations of the traditional PSO approach.

Scatter Search V1 (STS) (FS12): STS method uses an attributes subset for scatter search through the attribute space. However, it is starting with many significant attributes subsets and stops when the output is higher than a specified threshold or no further chances of improvement.

4.6 ML classifiers

In this study, five ML classifiers are used for designing CPDP such as Stochastic Gradient Descent (SGD), Sequential Minimal Optimization (SMO), AdaBoost (AB), Bagging (BNG), and LogiBoost (LB).

4.7 Statistical test

In this study, Friedman and Wilcoxon signed rank test (Liu and Xu 2022) are used for validating the difference among the performance of various FS techniques for CPDP. The results are validated using Friedman test and Wilcoxon sign rank test. However, the Wilcoxon sign rank test is used as a post hoc test in this study.

5 Results

The results and analysis are discussed in this section. We have collected eight projects of AEEEM and two projects of ReLink for experimentation. Furthermore, for experimenting all eight projects are combined to form different pairs for CPDP. However, these pairs are formed considering the size of the training dataset is low. Moreover, the model is not more efficient, if the amount of training data is less than the testing data. In the initial stage, 26 pairs are formed in this study, after extracting only those pairs having source data larger than the target data 13 pairs are selected. Furthermore, 12 methods are used for FS categorized as five filter methods, one wrapper subset method, and six swarm search-based methods. In this study, the FS techniques removed those features which are least relevant for the prediction of defects in the upcoming projects. In filter techniques, the features are ranked based on their contribution for defect prediction. Based on the number of features required for the prediction model, the relevant features are selected, and the remaining features are discarded. In the case of wrapper-embedded methods for FS, the cost function is optimized to find out the relevant features from both the source and target projects. Moreover, the features based on methods, attributes, and entropy are selected in the majority cases. Furthermore, five ML classifiers are used for the development of a prediction model with selected features.

In this study, 780 models were designed considering possible pairs of source and target datasets. The performance of each prediction model is analyzed according to the ML classifier for every attribute selection method. Each ML classifier takes a varied amount of time for the prediction model. The time complexity (Chen et al. 2020) of the SGD classifier is $O(r)$ where r is a number of records in the dataset, SMO classifier (Ouyang and Gray 2010) is $O(r^{2.3})$, BNG classifier (Zhao et al. 2020) is $O(Tt)$ where T refers to the number of

iterations and t is the average running time of an individual ML classifier, AB is $O(rlogr)$. However, the performance metric used in this study is AUC. AUC provides more generalized and unbiased results in case of an imbalanced dataset. This section interprets the results and answers to the RQs discussed in Sect. 1.

The existing FS algorithm is combined with a two-phase ensemble learning framework to improve the performance of software defect prediction models using SVM, random forest, and Enhanced Average Probability Ensemble (EAPE) technique. EAPE provided better results close to 0.9 for some of the datasets in comparison to SVM and random forest which are close to 0.7 and 0.8. Moreover, the specified FS techniques are used in the existing study using the same project for training and testing. The authors have used feature learning for CPDP using AEEEM and ReLink datasets with AUC values of 0.8 and 0.7 (Wu et al. 2018).

In the existing studies, authors have not analyzed the capability of filter, wrapper, and swarm search-based methods for FS. In the existing study, authors experimented with GA for CPDP. However, it is concluded that GA outperformed FS in comparison to the filtering technique (Hu and Zhu 2023). The authors also experimented with search-based FS techniques for CPDP with a KNN filter. Furthermore, the CPDP model using search based optimizer provided 30–40% better performance in comparison to the base paper. In this study, the performance of all the search-based FS techniques was analyzed with filter and wrapper techniques. However, it has been observed that search-based techniques outperformed for CPDP in comparison with the ranking and wrapper method. In the existing study, the FS methods are empirically validated for CPDP (Yu et al. 2019). The experiment is conducted using KNN and NB classifiers using feature subset selection and filtering techniques.

RQ1: What is the predictive capability of filter methods for CPDP using AEEEM and ReLink dataset?

In this study, CPDP models are developed successfully in conjunction with software metrics. Various filter methods are used for a selection of relevant attributes. Thus, filter methods are used with a ranker to select relevant features based on ranking. Null Hypothesis (H_o): There is no significant difference among the performance of filter methods used for FS. Alternate Hypothesis (H_a): There is a significant difference among the performance of filter methods used for FS. The predictive performance of filter methods was analyzed using the Friedman test for 325 CPDP models. Furthermore, the results obtained using filter methods are validated using Friedman test (Ma et al. 2023) at a significance level of 0.05. However, for five FS techniques and Degree of Freedom (DF) as 4, the obtained chi-square (χ^2) value is 7.520, and the p-value is 0.111. Moreover, the p-value is greater than the significance level which is 0.05. Thus, hypothesis testing proved that there is no significant difference among the performance of filter methods for five ML classifiers for the identification of defects in the CPDP. Hence, H_o is accepted and stated that filter methods performed similarly for FS with respect to HoGCP. The results are presented in Table 4.

RQ2: What is the predictive capability of swarm search-based methods for CPDP?

To answer this RQ, various swarm search-based methods are used for the selection of relevant attributes in this study. Thus, filter methods are used with correlation-based FS (CFS) to select relevant features. Null Hypothesis (H_o): There is no significant difference among the performance of six swarm search methods used for FS. Alternate Hypothesis (H_a): There is a significant difference among the performance of six swarm search methods used for FS. The results are presented in Table 5.

The predictive performance of swarm search-based methods analyzed using the Friedman test for 390 CPDP models. Furthermore, the results obtained using filter methods are validated using the Friedman test at a significance level of 0.05. However, for five FS techniques and DF as 4, the obtained χ^2 value is 7.353, and p-value is 0.196. Moreover, the p-value is greater than the significance level which is 0.05. Thus, hypothesis testing proved that there is no significant difference among the performance of six swarm search-based methods used with five ML classifiers for CPDP. Hence, H_o is accepted and stated that all the swarm search methods performed similarly for FS with respect to HoGCP.

RQ3: What is the predictive capability of filter, wrapper, and swarm search-based methods for CPDP?

To answer this RQ, the performance of the prediction model build using filter, wrapper, and swarm search-based methods is analyzed. Furthermore, the performance of 780 CPDP models are compared altogether using the Friedman test at a significance level of 0.05. H_o: There is no significant

Table 4 Average AUC values for filter methods and ML classifiers for CPDP (13 pairs of AEEEM and ReLink dataset)

ML classifier	Filter method				
	FS1	FS2	FS3	FS4	FS5
SGD	0.59	0.59	0.54	0.60	0.54
SMO	0.61	0.61	0.60	0.63	0.58
AB	0.75	0.75	0.75	0.68	0.65
BNG	0.64	0.64	0.66	0.62	0.63
LB	0.72	0.71	0.72	0.67	0.63

Table 5 Average AUC values for swam search-based methods and ML classifiers for CPDP (13 pairs of AEEEM and ReLink dataset)

ML classifier	Swarm search method					
	FS7	FS8	FS9	FS10	FS11	FS12
SGD	0.58	0.55	0.58	0.59	0.58	0.59
SMO	0.62	0.61	0.62	0.64	0.61	0.61
AB	0.73	0.74	0.73	0.73	0.73	0.74
BNG	0.7	0.72	0.7	0.69	0.69	0.70
LB	0.75	0.73	0.75	0.72	0.69	0.76

difference among the performance of filter, wrapper, and swarm search methods used for FS. H_a: There is a significant difference among the performance of filter, wrapper, and swarm search methods used for FS. However, for 12 FS techniques and DF as 11, the obtained χ^2 value is 23.185, and p-value is 0.017. However, the p-value is less than the significance level which is 0.05. Furthermore, the Friedman test resulted in the acceptance of an H_a, that is there is a significant difference among the performance of all three methods and 12 techniques used. Thus, H_a is accepted and stated that the performance of filter, wrapper, and swarm search-based FS methods differ significantly for HoGCP.

Furthermore, the results are validated using post hoc analysis with the Wilcoxon sign rank test. Wilcoxon sign rank test performed pairwise comparison among all the 12 techniques that formed 66 combinational pairs. Furthermore, out of 66 pairs, the Wilcoxon sign rank test resulted in the acceptance of the null hypothesis for 18 pairs presented in Table 6. Thus, there is no significant difference among the performance of 18 pairs of FS techniques. However, for the remaining 48 pairs Wilcoxon sign rank tests resulted in the acceptance of the alternate hypothesis presented in Table 7. Hence, there is a significant difference among the performance of 48 combinational pairs of FS techniques. Thus, the performance of swarm search-based methods is comparatively better than filter and wrapper methods.

Table 6 Wilcoxon signed-rank test result for analyzing the performance of filter, wrapper, and swarm search-based feature selection methods (18 pairs not significant)

Filter, Wrapper, and Swarm Search-based FS Pairs	Asymp. Sig. (2-tail)
FS5–FS2	0.043
FS5–FS3	0.039
FS7–FS5	0.042
FS8–FS5	0.042
FS9–FS5	0.042
FS10–FS5	0.042
FS11–FS5	0.041
FS12–FS5	0.043
FS7–FS6	0.038
FS8–FS6	0.041
FS9–FS6	0.038
FS10–FS6	0.042
FS11–FS6	0.042
FS12–FS6	0.042
FS9–FS7	0.025
FS11–FS7	0.042
FS11–FS9	0.042
FS12–FS11	0.042

6 Conclusion

The objective of conducting this study was to analyze the effectiveness of FS techniques for CPDP. However, in this study, the HoGCP model is designed, that considered source and target projects with similar features. In this study, three types of methods are explored for CPDP (i) Filter method (ii) Wrapper subset evaluation method, and (iii) Swarm search-based methods. Also, this study is evaluated using five ML classifiers for defect prediction in future projects with similar features to enhance the maintainability, reliability, and robustness of the software. Friedman and Wilcoxon sign rank test were used for statistical validation.

Five filter methods are compared based on ranking criteria. However, all the filter methods were performed on the same scale. However, for wrapper attribute subset evaluation, the principle of greediness is used to find out the global optimal subset of attributes with a specified ML classifier. However, the wrapper method is better than the filter method in most of the pairs. Furthermore, six swarm search-based methods are used for FS. However, the swarm method finds out the best attribute through attributes pairs search space and stops selecting the attribute according to the stopping criterion. Furthermore, swarm search-based methods outperformed with respect to filter and wrapper methods. Hence, conducting further experiments for CPDP swarm search-based methods are preferable. Furthermore, the FS techniques need to be explored for HeGCP defect prediction models on large datasets.

Table 7 Wilcoxon signed-rank test result for analyzing the performance of filter, wrapper, and swarm search-based feature selection methods (48 pairs not significant)

Filter, Wrapper, and Swarm Search-based FS Pairs	Asymp. Sig. (2-tail)	Filter, Wrapper, and Swarm Search-based FS Pairs	Asymp. Sig. (2-tail)
FS2–FS1	0.686	FS9–FS3	0.078
FS3–FS1	0.892	FS10–FS3	0.279
FS4–FS1	0.180	FS11–FS3	1.000
FS5–FS1	0.129	FS12–FS3	0.078
FS6–FS1	0.684	FS5–FS4	0.129
FS7–FS1	0.225	FS6–FS4	0.684
FS8–FS1	0.498	FS7–FS4	0.345
FS9–FS1	0.225	FS8–FS4	0.892
FS10–FS1	0.080	FS9–FS4	0.225
FS11–FS1	1.000	FS10–FS4	0.176
FS12–FS1	0.225	FS11–FS4	1.000
FS3–FS2	0.336	FS12–FS4	0.225
FS4–FS2	0.892	FS6–FS5	0.098
FS6–FS2	0.102	FS8–FS7	0.498
FS7–FS2	0.414	FS10–FS7	0.680
FS8–FS2	1.000	FS12–FS7	0.157
FS9–FS2	0.223	FS9–FS8	0.276
FS10–FS2	0.500	FS10–FS8	0.686
FS11–FS2	0.343	FS11–FS8	0.223
FS12–FS2	0.225	FS12–FS8	0.104
FS4–FS3	0.890	FS10–FS9	0.396
FS6–FS3	0.102	FS12–FS9	0.492
FS7–FS3	0.276	FS11–FS10	0.223
FS8–FS3	0.496	FS12–FS10	0.684

Funding This research is not supported under any funding.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Research involving human participants and/or animals This is an experimental study. This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

References

Akman DV, Malekipirbazari M, Yenice ZD, Yeo A, Adhikari N, Wong YK, Abbasi B, Gumus AT (2023) k-best feature selection and ranking via stochastic approximation. Expert Syst Appl 213:118864. https://doi.org/10.1016/j.eswa.2022.118864

Bachmann A, Bird C, Rahman F, Devanbu P, Bernstein A (2010) The missing links: bugs and bug-fix commits. In: Proc ACM SIGSOFT Symp Found Softw Eng, pp 97–106 https://doi.org/10.1145/1882291.1882308

Bai L, Wang Z, Shao YH, Deng NY (2014) A novel feature selection method for twin support vector machine. Knowledge-Based Syst 59:1–8. https://doi.org/10.1016/j.knosys.2014.01.025

Bhutamapuram US, Sadam R (2022) With-in-project defect prediction using bootstrap aggregation based diverse ensemble learning technique. J King Saud Univ - Comput Inf Sci 34:8675–8691. https://doi.org/10.1016/j.jksuci.2021.09.010

Bird C, Bachmann A, Aune E, Duffy J, Bernstein A, Filkov V, Devanbu P (2009) Fair and balanced? Bias in bug-fix datasets. In: ESEC-FSE'09 - Proc Jt 12th Eur Softw Eng Conf 17th ACM SIGSOFT Symp Found Softw Eng, pp 121–130. https://doi.org/10.1145/1595696.1595716

Briand LC, Melo WL, Wüst J (2002) Assessing the applicability of fault-proneness models across object-oriented software projects. IEEE Trans Softw Eng 28:706–720. https://doi.org/10.1109/TSE.2002.1019484

Chen X, Lee JD, Tong XT, Zhang Y (2020) Statistical inference for model parameters in stochastic gradient descent. Ann Stat 48:251–273. https://doi.org/10.1214/18-AOS1801

Czibula G, Marian Z, Czibula IG (2014) Software defect prediction using relational association rule mining. Inf Sci (ny) 264:260–278. https://doi.org/10.1016/j.ins.2013.12.031

Faiz RB, Shaheen S, Sharaf M, Rauf HT (2023) Optimal feature selection through search-based optimizer in cross project. Electronics 12:514. https://doi.org/10.3390/electronics12030514

Feng F, Li K-C, Yang E, Zhou Q, Han L, Hussain A, Cai M (2023) A novel oversampling and feature selection hybrid algorithm for imbalanced data classification. Multimed Tools Appl 82:3231–3267. https://doi.org/10.1007/s11042-022-13240-0

Hawkins DM (2004) The problem of overfitting. J Chem Inf Comput Sci 44:1–12. https://doi.org/10.1021/ci0342472

He Z, Peters F, Menzies T, Yang Y (2013) Learning from open-source projects: an empirical study on defect prediction. Int Symp Empir Softw Eng Meas. https://doi.org/10.1109/ESEM.2013.20

Herbold S (2016) CrossPare: a tool for benchmarking cross-project defect predictions. In: Proc. - 2015 30th IEEE/ACM Int. Conf. Autom. Softw. Eng. Work. ASEW 2015, pp 90–95. https://doi.org/10.1109/ASEW.2015.8

Hu Z, Zhu Y (2023) Cross-project defect prediction method based on genetic algorithm feature selection. Eng Reports. https://doi.org/10.1002/eng2.12670

Khatri Y, Kumar SS (2023) An effective feature selection based cross-project defect prediction model for software quality improvement. Int J Syst Assur Eng Manag 14:154–172. https://doi.org/10.1007/s13198-022-01831-x

Kim S, Zhang H, Wu R, Gong L (2011) Dealing with noise in defect prediction. In: Proc - Int Conf Softw Eng, pp 481–490. https://doi.org/10.1145/1985793.1985859

Laradji IH, Alshayeb M, Ghouti L (2015) Software defect prediction using ensemble learning on selected features. Inf Softw Technol 58:388–402. https://doi.org/10.1016/j.infsof.2014.07.005

Lessmann S, Baesens B, Mues C, Pietsch S (2008) Benchmarking classification models for software defect prediction: a proposed framework and novel findings. In: IEEE transactions on software engineering, pp 485–496. https://doi.org/10.1109/TSE.2008.35

Li Y, Li T, Liu H (2017) Recent advances in feature selection and its applications. Knowl Inf Syst 53:551–577. https://doi.org/10.1007/s10115-017-1059-8

Liu J, Xu Y (2022) T-Friedman test: a new statistical test for multiple comparison with an adjustable conservativeness measure. Int J Comput Intell Syst 15:1–19. https://doi.org/10.1007/s44196-022-00083-8

Liu H, Yu L (2005) Toward integrating feature selection algorithms for classification and clustering. IEEE Trans Knowl Data Eng 17:491–502. https://doi.org/10.1109/TKDE.2005.66

Liu H, National U of S, Motoda H, Osaka U, Setiono R, Zhao Z (2010) Feature Selection : an ever evolving frontier in data mining. J Mach Learn Res Work Conf Proc 10 Fourth Work Featur Sel Data Min, pp 4–13

Liu J, Ai J, Lu M, Wang J, Shi H (2023) Semantic feature learning for software defect prediction from source code and external knowledge. J Syst Softw 204:111753. https://doi.org/10.1016/j.jss.2023.111753

Lyu Y, Feng Y, Sakurai K (2023) A survey on feature selection techniques based on filtering methods for cyber attack detection. Information 14:191. https://doi.org/10.3390/info14030191

Ma Y, Luo G, Zeng X, Chen A (2012) Transfer learning for cross-company software defect prediction. Inf Softw Technol 54:248–256. https://doi.org/10.1016/j.infsof.2011.09.007

Ma Z, Wu G, Suganthan PN, Song A, Luo Q (2023) Performance assessment and exhaustive listing of 500+ nature-inspired metaheuristic algorithms. Swarm Evol Comput 77:101248. https://doi.org/10.1016/j.swevo.2023.101248

Malhotra R, Meena S (2022) Defect prediction model using transfer learning. Soft Comput 26:4713–4726. https://doi.org/10.1007/s00500-022-06846-x

Nam J, Pan SJ, Kim S (2013) Transfer defect learning. In: Proceedings - International Conference on Software Engineering, pp 382–391. IEEE Computer Society https://doi.org/10.1109/ICSE.2013.6606584

Ouyang H, Gray A (2010) Fast stochastic frank-wolfe algorithms for nonlinear SVMs. In: Proc. 10th SIAM Int. Conf. Data Mining, SDM 2010, pp 245–256. https://doi.org/10.1137/1.9781611972801.22

Pal S, Sillitti A (2022) Cross-project defect prediction: a literature review. IEEE Access 10:1–1. https://doi.org/10.1109/access.2022.3221184

Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/TKDE.2009.191

Pramanik R, Pramanik P, Sarkar R (2023) Breast cancer detection in thermograms using a hybrid of GA and GWO based deep feature selection method. Expert Syst Appl 219:119643. https://doi.org/10.1016/j.eswa.2023.119643

Rahman F, Posnett D, Devanbu P (2012) Recalling the "imprecision" of cross-project defect prediction. In: Proceedings of the ACM SIGSOFT 20th international symposium on the foundations of software engineering, FSE 2012. p 1. ACM Press, New York. https://doi.org/10.1145/2393596.2393669

Shaheen M, Naheed N, Ahsan A (2023) Relevance-diversity algorithm for feature selection and modified Bayes for prediction. Alexandria Eng J 66:329–342. https://doi.org/10.1016/j.aej.2022.11.002

Sharma T, Jatain A, Bhaskar S, Pabreja K (2023) Ensemble machine learning paradigms in software defect prediction. Procedia Comput Sci 218:199–209. https://doi.org/10.1016/j.procs.2023.01.002

Tang C, Zheng X, Zhang W, Xinwang L, Xinzhong Z, En Z (2023) Unsupervised feature selection via multiple graph fusion and feature weight learning. Sci China Inf Sci 66:152101. https://doi.org/10.1007/s11432-022-3579-1

Tantithamthavorn C, McIntosh S, Hassan AE, Matsumoto K (2017) An empirical comparison of model validation techniques for defect prediction models. IEEE Trans Softw Eng 43:1–18. https://doi.org/10.1109/TSE.2016.2584050

Tong H (2019) AEEEM-JIRA-PROMISE. https://zenodo.org/record/3362613

Tong H, Wei L, Weiwei X, Wang S (2023) ARRAY: adaptive triple feature-weighted transfer Naive Bayes for cross-project defect prediction. J Syst Softw 202:111721. https://doi.org/10.1016/j.jss.2023.111721

Vashisht R, Murtaza RSA (2023) Feature engineering to heterogeneous cross software projects defect prediction: a Novel framework. Arab J Sci Eng 48:2539–2560. https://doi.org/10.1007/s13369-022-07337-9

Vashisht R, Rizvi SAM (2020) Heterogeneous cross project defect prediction—a survey. In: Communications in computer and information science, pp 278–288, Springer. https://doi.org/10.1007/978-981-15-4451-4_22

Visalakshi S, Radha V (2015) A literature review of feature selection techniques and applications: review of feature selection in data mining. In: 2014 IEEE Int. Conf. Comput. Intell. Comput. Res. IEEE ICCIC 2014. https://doi.org/10.1109/ICCIC.2014.7238499

Wu F, Jing XY, Sun Y, Sun J, Huang L, Cui F, Sun Y (2018) Cross-project and within-project semisupervised software defect prediction: a unified approach. IEEE Trans Reliab 67:581–597. https://doi.org/10.1109/TR.2018.2804922

You G, Wang F, Ma Y (2016) An empirical study of ranking-oriented cross-project software defect prediction. Int J Softw Eng Knowl Eng 26:1511–1538. https://doi.org/10.1142/S0218194016400155

Yu Q, Jiang S, Qian J (2016) Which is more important for cross-project defect prediction: instance or feature? In: Proceedings - 2016 International Conference on Software Analysis, Testing and Evolution, SATE 2016, pp 90–95. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SATE.2016.22

Yu Q, Jiang S, Zhang Y (2017) A feature matching and transfer approach for cross-company defect prediction. J Syst Softw 132:366–378. https://doi.org/10.1016/j.jss.2017.06.070

Yu Q, Qian J, Jiang S, Wu Z, Zhang G (2019) An empirical study on the effectiveness of feature selection for cross-project defect prediction. IEEE Access 7:35710–35718. https://doi.org/10.1109/ACCESS.2019.2895614

Zhao Z, Karimzadeh M, Gerber F, Braun T (2020) Mobile crowd location prediction with hybrid features using ensemble learning. Futur Gener Comput Syst 110:556–571. https://doi.org/10.1016/j.future.2018.06.025

Zhao Y, Yi Z, Yu Q, Chen X (2022) Cross-project defect prediction considering multiple data distribution simultaneously. Symmetry (basel) 14:441

Zimmermann T, Nagappan N, Gall H, Giger E, Murphy B (2009) Cross-project defect prediction. 91. https://doi.org/10.1145/1595696.1595713

Zong X, Li G, Zheng S, Zou H, Yu H, Gao S (2023) Heterogeneous cross-project defect prediction via optimal transport. IEEE Access 11:12015–12030. https://doi.org/10.1109/ACCESS.2023.3241924

Zou J, Li Z, Liu X, Tong H (2023) MSCPDPLab: A MATLAB toolbox for transfer learning based multi-source cross-project defect prediction. SoftwareX 21:101286. https://doi.org/10.1016/j.softx.2022.101286

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

J. Water Environ. Nanotechnol., 8(3): 267-284 Summer 2023

ORIGINAL RESEARCH PAPER

Energy matrices and life cycle conversion analysis of N-identical hybrid double slope solar distiller unit using Al_2O_3 nanoparticle

Dharamveer Singh* [1,2], ***Ashok Kumar Yadav***[3], ***Anil Kumar***[4], ***Samsher***[4,5]

[1]*Department of Mechanical Engineering, R. D Engineering College, Ghaziabad, U.P., India*
[2]*Research Centre, M. R. D. Trust, Modinagar, Ghaziabad, U.P., India*
[3]*Department of Mechanical Engineering, Raj Kumar Goel in stitute of technology, Ghaziabad, U P. india*
[4]*Department of Mechanical Engineering, Delhi Technological University, Delhi, India*
[5]*Department of Mechanical Engineering, H. B. T. U., Kanpur, U. P., India*

Received: 2023-02-19 Accepted: 2023-04-24 Published: 2023-08-10

ABSTRACT

In the current study, 25% incorporating PVT hybrid CPC collector double slope solar still is using Al_2O_3 nanoparticles underwent energy matrices analysis and life cycle conversion efficiency (LCCE). With the aid of an analytical program fed into MATLAB, the analysis is conducted on an annual basis based on the atmospheric conditions in New Delhi. The IMD in Pune, India, provided weather input data needed for the numerical computations. The average annual energy output will be calculated using energy and exergy and later on evaluated EPT, EPF, and LCCE. This will reveal that the average annual yield is 8.5%, the average energy payback time is 16.16%, the average energy payback factor is 13.91%, and the average life cycle cost conversion efficiency is 7.15% higher compared to the previous research. Therefore, it is obvious the proposed system is better based on the following parameters i.e. (i). annual yield, (ii). energy matrices such as EPT, EPF, and efficiency of life cycle cost (LCCE). The proposed hybrid is a self-sustainable system that can also meet the future requirements of potable water as well as electricity.

Keywords: Energy matrices, Energy payback time, Energy production factor, Life cycle conversion efficiency, Al_2O_3 nanoparticles

How to cite this article

Singh D., Kumar Yadav A., Kumar A., Samsher., Energy matrices and life cycle conversion analysis of N-identical hybrid double slope solar distiller unit using Al_2O_3 nanoparticle". J. Water Environ. Nanotechnol., 2023; 8(3): 267-284.
DOI: 10.22090/jwent.2023.03.006

INTRODUCTION

We can access a very low percentage of water from the ground. Therefore, there is a need to develop potable water and self-sustainable systems. Water purification is required due to polluted water to freshwater throughout the world. Consumption of polluted water is increasing death rates by increasing diseases in human beings. In present days, system availability is not self-sustainable. Electricity is needed which generates it causes pollution. Therefore, the better solution is a renewable energy source that can reduce the potable water problem. Lawrence and Tiwari [1] developed the empirical relations for the inside coefficients of heat transfer from the natural flow with a heat exchanger in a solar distiller unit. Popiel and Wojtkowiak [2] studied the thermo-physical properties of the base fluid. Pak and Cho [3] evaluated various correlations for different properties. G. N. Tiwari [4] studied the fundamental design of the solar still. Hwang et al. [5] analyzed the heat transfer coefficient for Al_2O_3 nanofluids.

* Corresponding Authors Email: *veerdharam76gmail.com*

This work is licensed under the Creative Commons Attribution 4.0 International License.
To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Barden [6] improved the thermo-physical properties of the base fluid; the heat transfer coefficients could also be improved. Due to their superior thermo-physical characteristics, nanoparticles are easily deferred. The nanofluids are developing fluids with extremely quick heat transfer properties. Additionally, the base fluid's qualities could be enhanced by customizing the size and shape. Tiwari and Tiwari [7] expressed few merits of solar distillers over other distillation technologies such as filters, membranes, and batteries, no definitive resource of energy, and primarily low investment. Ho et al. [8] numerically analyzed nanofluids for natural convection in a square enclosure: effects due to uncertainties of viscosity and thermal conductivity. Otanicar and Golden [9] analyzed the enviro economic aspect of solar collectors using nanofluid and found it neutralizes 74 kg for a life span of 15 years. Patel et al. [10] found the thermal conductivity of nanofluids. Singh et al. [11] theoretically investigated entropy generation for nanofluids. Elzen et al. [12] analyzed emission reductions, abatement costs, and carbon prices. Khanafer and Vafai [13] This work presented the thermophysical properties of nanofluids. Khullar and Tyagi [14] analyzed and reported emissions of 103 kg approx./ household/year reduced for a solar heating device for nanofluids. Faizel et al. [15] analyzed based on the cost of flat plate collector (FPC) using tin oxide, copper oxide, titanium oxide, and aluminum oxide) nanofluids. It is discovered that the high density, low specific heat, and thermal conductivity of CuO nanofluid are more appropriately attributed to its performance.Liu et al. [16] have evaluated the economic analysis of the integrated solar distiller unit of the evacuated tube. Kabeel et al. [17] analyzed the sole inclined solar distiller unit with vacuum as a water-based nanofluid. Elango et al. [18] analyzed practically single-slope solar distillers as thermal energy, exergy, and productivity using different nanofluids. Omara et al. [19] analyzed the performance of corrugated wick type and simple solar distiller units using nanofluids. Tiwari et al. [20] analyzed experimentally active solar distillers that exergoeconomic and environmental economic using water-based nanofluid the photovoltaic thermal flat plate collector is met potable water requirements daily. Environmental damage has been estimated to cost $6.29 annually. Sharon and Reddy [21] analyzed the annual economic performance of an active solar distiller loaded with saline water. Sahota et al. [22, 24] analyzed the passive double slope solar distiller unit performance using nanofluids and concluded that the aluminum oxide-based nanofluid gives better performance than others. Singh et al. [23] analyzed the energy matrix and existence cycle conversion efficiency for conventional single and double slope distiller units and found 0.144. and 0.137 per unit cost, respectively, and exergoeconomic parameters. Singh and Tiwari [25] analyzed the energy matrices and life cycle cost of an active partly PVT-CPC solar distiller. Shashir et al. [26] analyzed the performance of nanoparticles like copper oxide and graphite micro-flakes on solar distiller units with different cooling on the cover of toughened glass. It is concluded the solar yield increases and copper oxide 47.8% and 57.6%. Sahota et al. [27] studied the performance of PVT-FPC double slope solar distiller unit with or without helical coil heat exchanger using nanofluid and found water-based nanofluid performance was better with a heat exchanger. Saleha et al. [28] analyzed the effect of solvent and found it effective in solar distiller units. Chen et al. [29] analyzed that experimentally found the stability of weak luminous was very good with nanofluid in solar distiller unit and the effect of brackish water's constancy, ocular and thermal properties using nanofluid feasible. Mahian et al. [30] studied and found a significant temperature lower than 50 °C, in a heat exchanger and found a 2 times greater amount of water than without a heat exchanger. Additionally, water nanoparticles improve evaporation at low temperatures. It is crucial to assess the cost-effectiveness of renewable energy systems based on payback. Sahota et al. [31] analyzed environmental economics and exergoeconomics for passive double slope solar distiller with water-loaded nanofluid (CuO,Al_2O_3,TiO_2) and found payback time of energy of the system is low and the cost of environmental per annum is higher on mitigation with nanofluid. Singh and Tiwari [32] analyzed the augmentation in energy matrices of N-PVT-FPC partly double slope solar distiller. Joshi and Tiwari[33] analyzed single slope N^{th}-identical photovoltaic thermal compound parabolic concentrator collector N-PVT-CPC. Dharamveer et al. [34] reviewed nanofluid-loaded desalination**.** Kumar and Singh [35] analyzed the Energy and exergy of active solar stills using a compound parabolic concentrator. Shanker, et al.[36] analyzed the performance of the C.I. engine using biodiesel fuel by modifying

injection timing and injection pressure Anup et al. [37] analyzed using FEA of refrigerator compartment for optimizing thermal efficiency. Kumar and Singh.[38]optimized thermal behavior of a small heat exchanger. Zhang et al. [39] presented in the area of sustainable energies focuses on utilizing green and clean technologies. Dhivagar et al. [40] analyzed single slope grate crude shrewd solar distiller units for energy, exergy, and economic aspects. Dharamveer and Samsher [41] studied the active and passive solar still behavior on energy matrices and enviroeconomics. Arora et al. [42] analyzed double slope solar distiller N-PVT-CPC using carbon nanotubes for water generation. Dharamveer et al. [43] analytically studied N^{th} identical photovoltaic thermal (PVT) compound parabolic concentrator (CPC) active double slope solar distiller with a helically coiled heat exchanger using C_uO Nanoparticles. Dharamveer, et al. [44] analyzed an N-identical active single-slope solar distiller with a helically coiled heat exchanger using CuO nanoparticles. Kumar and Singh, [45] compared single-phase microchannels for heat flow Experimental and using CFD. Subrit and Singh. [46] analyzed thermal of coal and waste cotton oil liquid produced by pyrolysis of diesel engine fuel was carried out by. Shahsavar et al. [47] Compared energy, exergy, environmental, exergoeconomic, and enviro economic analysis of building integrated photovoltaic/thermal, earth-air heat exchanger, and hybrid systems. Numerous studies on passive and active solar stills have been conducted, according to the current literature survey. However, not much investigation on active solar still filled with water-based nanofluids was studied. Based on energy and exergy, Dharamveer et al. studied a hybrid double slope. No scholars have examined the economic, and environmental, using nanofluid. Furthermore, no studies have been conducted for CPC, ETC double slope basin type solar distiller using nanofluid. Thus, the proposed study will examine the impacts of active solar still double slope with CPC and filled with water containing Al_2O_3 nanofluid on energy matrices based such as energy payback time, energy payback factor, life cycle cost conversion efficiency, and productivity, of solar desalination systems will be thoroughly examined. The effectiveness of the suggested approach will also be evaluated in comparison to the findings of past studies.

MATERIALS AND METHODS

To determine the following objectives the methodology is adopted as energy matrices analysis of hybrid solar distiller basin type double slope with heat exchanger using Al_2O_3 nanoparticles.

System Description

Working of double slope solar desalination incorporating PVT with CPC collector using nanoparticles (N-PVT-CPC-DS-HE) is shown. Representation of solar still is followed the greenhouse principle. The parameters used for the distiller unit are given in Table 1 and Table 2. The basin of the solar distiller is connected in series with N-CPC and incorporated with a helically coiled heat exchanger. Solar radiation received on the glass cover is transferred to the water surface and thereafter absorbed by the surface. Later on, reflected water received heat on the top cover and rest portion and then move to the liner where the maximum amount absorbed and liner temperature increased and transfer this heat to water. Thus the temperature of water in the basin is increased and water gets evaporated. Collectors also heat the water in the basin. Water is heated and evaporates in this manner. The distillate trickles forward to the passage attached to the bottom side as the vapor reaches the interior face of the condenser, where film-wise condensation takes place. Then, the beaker receives the distilled water that was siphoned off.

Fig. 1 represents 25% PVT incorporating hybrid solar still. Collectors are put south facing at an angle of 45° which are connected in series as the input of the second collector is attached to the first collector output. Radiation that falls on the collector directly gets absorbed and beam irradiation is reflected on the parabolic concentrator. Similarly, irradiation falls on PV modules that generate electricity which is further used for operating pumps (D.C) and access energy can be further used for any electrical appliances according to need. Table 2 represents the specifications of the proposed system. The inclination angle of the system is 30° and the orientation of the system is the southern face. The basin liner is absorbed maximum radiation which falls on the glass cover. The basin fluid gets heat from the heat exchanger through Al_2O_3 nanoparticles. As per a prior study, it is obvious that a helically coiled heat exchanger is more effective than any other design. The Al_2O_3 nanoparticles exchange more heat in active solar still because it covers more

Table 1. Specifications of double slope PVT-CPC-DS active solar distillation system [43]

Hybrid solar distiller unit			
Components		Particulars	
Basin length		2.0 m	
Basin width		1.0 m	
Tilting glass cover		15°	
Basin (smaller side) height		0.2 m	
Body's material		G.R.P	
Stand's material		G.I.	
Covers		Glass	
Collector facing		south	
Tilting glass thickness		0.004 m	
K_g		0.816 W/mK	
Insulation width		0.1 m	
Thermal conductivity of insulation		0.166 W/mK	
Hybrid collector			
Component's name	Specifications	Component's name	Specifications
Kinds of collector	Tube kinds	Area of aperture	2 m^2
Receiver area	1.0 m^2	Aperture's module area	1.0 m^2
Collector's plate thick	0.0020 m	Aperture's receiver area	1.5 m^2
Copper's tube thick	0.00056 m	Module's receiver area	0.25 m^2
Copper's tube length	1.0 m	Collector's receiver area	0.75 m^2
Packing coefficient	0.8	No of collectors	N=4
D.C. motor	12V, 40 Watt	Pipe dia.	0.0125 m
Inclination of CPC	30°	Inclination of FPC	30°
Under glass effective area	0.75 m^2	Under PV module effective area	0.660 m^2
No. of the cell (solar)	36	Single-cell (Solar)	0.007 m^2
Basin area	2 m^2	m_f	0.02 kg/s
Heat exchanger (copper coiled)			
No. of turns	12	Tube coil dia	0.0125 m
Length of heat exchanger	1.937 m	Heat exchanger coil dia	0.045 m

Table 2. Parameters are used for calculations [43]

Specifications	Numerical values	Specifications	Numerical values
α_g	0.050	A_{am}	0.5
α_b	0.5861	A_{ac}	1.5
α_{bf}	0.82	L_p	0.0020
α_c	0.90	K_m	64 W/mK
α_p	0.80	K_p	64 W/mK
β_c	0.89	L_i	0.1m
K_g	0.8160 W/mK	h_i	5.7 W/m^{2}K
K_b	0.035 W/mK	h_0	9.5 W/m^{2}K
K_p	0.166 W/mK	U_{tcp}	5.5451 W/m^{2}K
L_g	0.004	U_{tca}	15.03 W/m^{2}K
L_c	0.005	U_{tpa}	5.56 W/m^{2}K
L_i	0.1	U_{Lm}	9.03 W/m^{2}K
β_0	0.0045/K	U_{LC}	5.43 W/m^{2}K
x	0.33m	PF_1	0.2695
σ	$5.67*10^{-8}$ (W/m^2K^4)	PF_2	0.9398
τ_g	0.95	PF_c	0.977
$F^/$	0.968	ϵ_g	0.95
η_0	0.15	ϵ_{bf}	0.95

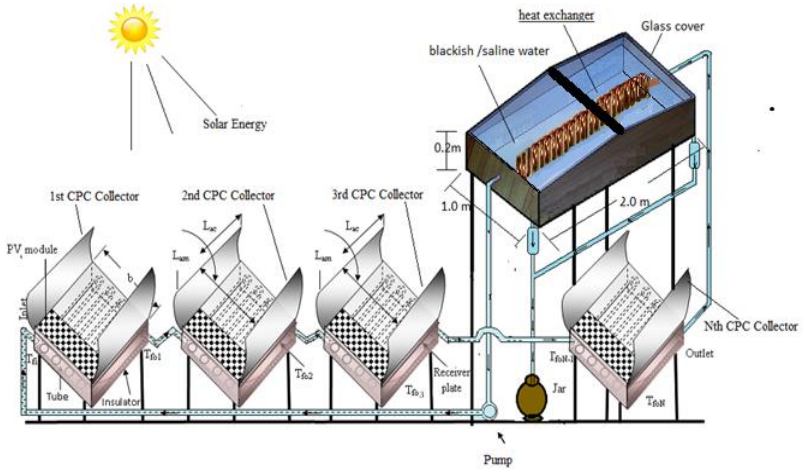


Fig. 1 active double slope solar still 25% incorporating N-PVT-CPC-DS-HE

surface area due to increasing volume due to heat exchanger. The system gets thermal energy via a combination of double-slope basins incorporating with PVT-CPC collector unit and it absorbs heat externally from a collector and internally from a basin. Basin water temperature increases via heat exchanger nanoparticles. Ultimately by releasing latent heat the vapor gets condensed, and collected at the lower end of the inclined glass cover of the basin.

This sort of system produced electricity as well as potable water. The other quality of this sort of system is i.e. low maintenance, easy to install, and useful for large and small demands of potable water as well as industrial purpose. Lot of advantages of such a system but in this work, we emphasize producing potable water only therefore according to making the system self-sustainable the PVT is provided otherwise 50%, 75%, and 100% can also be used.

Though, the proposed system-A is compared with the previous System-B based on basin water temperature, inside glass cover temperature, water outlet temperature, the overall thermal energy of collectors, overall exergy, overall electrical exergy, and potable water amount. Sedimentation possibility in nanofluid is more. Therefore nanoparticles size leads to a change in aggregation. The value of size matters to change in aggregation. Later on, more sophisticated equipment is needed to remove it.

Governing equations

To develop the characteristic equation, the following assumptions are:

i. Constant water level
ii. Negelectedohmic losses
iii. No leakage
iv. Over the entire surface film condensation
v. Steady-state partially covered active solar stills

The governing equations of the system are as follows:

a. East face

$$\alpha_g I_{SE} A_{gE} + h_{1wE}(T_w - T_{giE})\frac{A_b}{2} - h_{EW}(T_{giE} - T_{giW})A_{gE} = U_{cgaE}(T_{giE} - T_a)A_l \quad (1)$$

b. West face

$$\alpha_g I_{SW} A_{gW} + h_{1wW}(T_w - T_{giW})\frac{A_b}{2} + h_{EW}(T_{giE} - T_{giW})A_{gW} = U_{cgaW}(T_{giW} - T_a)A_{gW} \quad (2)$$

On solving Equations (1) and (2)

$$T_{giE} = \frac{A_1 + A_2 T_w}{p} \quad (3)$$

$$T_{giW} = \frac{B_1 + B_2 T_w}{p} \quad (4)$$

$$T_{goE} = \frac{\frac{K_g}{l_g} T_{glE} + h_{1gE} T_a}{\frac{K_g}{l_g} + h_{1gE}} \quad (5)$$

$$T_{goW} = \frac{\frac{K_g}{l_g} T_{glW} + h_{1gW} T_a}{\frac{K_g}{l_g} + h_{1gW}} \quad (6)$$

The unknown terms, A_1, A_2, B_1, B_2, and P in Equations 3, 4, 5, and 6 are mentioned in Appendix A,

c. Basin liner

$$\alpha_b(I_{SE} + I_{SW}) + 2h_{bw}(T_b - T_w) + 2h_{ba}(T_b - T_a) \quad (7)$$

d. Water equation

$$m_f C_f \frac{dT_w}{d_t} = \alpha_w(I_{SE} + I_{SW})\frac{A_b}{2} + 2h_{bw}(T_b - T_w)\frac{A_b}{2} - h_{1wE}(T_w - T_{giE})\frac{A_b}{2} - h_{1wW}(T_w - T_{giW})\frac{A_b}{2} + Q_{uN} \quad (8)$$

e. Energy balance equations for heat exchange and base fluid can be written for solar distillers as:

$$m_f C_f \frac{dT_w}{d_x} d_x = -(2\pi r_{11} U)(T_{HE} - T_w)d_x \quad (9)$$

Applying the limits Tw at (x = 0) = T_{woN} and Tw at (x = L) = T_{wi} on solving,

$$T_{wi} = T_{HE}\left[1 - \exp\left(\frac{-(2\pi r_{11} UL)}{m_f C_f}\right)\right] + T_{woN}\exp\left(\frac{-(2\pi r_{11} UL)}{m_f C_f}\right) \quad (10)$$

Where $U = \left[\frac{1}{h_{bf}} + \left(\frac{r_{11}}{k_1}\right)\ln\left(\frac{r_{22}}{k_1}\right)\left(\frac{1}{h_{bf}}\right)\right]^{-1}$

f. Energy balance for water collector N-PVT-CPC

$$T_{woN} = \left[\frac{(AF_R(\alpha\tau))_1(1 - K_p^N)}{m_f C_f(1 - K_p)}\right] I_b + \left[\frac{(AF_R(UL)_1)(1 - K_p^N)}{m_f C_f(1 - K_p)}\right] T_a + T_{wi} K_m^N \quad (11)$$

$$T_{woN} = \left[\left[\frac{(AF_R(\alpha\tau))1(1 - K_p^N)}{m_f C_f(1 - K_p)}\left(\frac{1}{(1 - e^z K_m^N)}\right)\right]\right] I_b +$$

$$\left[\frac{(AF_R(UL)1)(1-K_p^N)}{m_f C_f(1-K_p)}\left(\frac{1}{(1-e^z K_m^N)}\right)\right]T_a + T_{HE}\left(\frac{(1-e^z)K_m^N}{(1-e^z K_m^N)}\right)\Bigg] \quad (12)$$

Using this relation heat gain computed

$$Q_{uN} = m_f C_f (T_{woN} - T_{wi}) \quad (13)$$

The outlet temperature of water computed by these two equations 3.12 and 3.13,

$$T_{HE} > T_{wi} > T_w$$

$$Q_{uN} = [((\frac{(AF_R(\alpha\tau))1\,(1-K_p^N)}{(1-K_p)})\left(\frac{1}{(1-e^z K_m^N)}\right))I_b + \left(\left(\frac{(AF_R(UL)1)(1-K_p^N)}{(1-K_p)}\right)\left(\frac{1}{(1-e^z K_m^N)}\right)\right)T_a + m_f C_f (T_{HE}(\frac{(1-e^z)K_m^N}{(1-e^z K_m^N)}) - T_{wi})] \quad (14)$$

By putting

T_{giE}, T_{giW}, $2h_{bw}$ (T_b-T_w), and Q_{uN} from Equations (3),(4), and (9) in Equation (15)

$$\frac{dT_w}{dt} = -a_2 T_w + f_2(t) \quad (15)$$

The energy and exergy analysis has been done, respectively, based on the thermodynamics laws for the given entities.

$$E_{hourlyEn} = [h_{1wE}(T_w - T_{giE}) + h_{1wW}(T_w - T_{giW})](A_b) \quad (16)$$

$$E_{hourlyEx} = \{h_{1wE}[(T_w - T_{g\,iE}) - (T_a + 273)\ln\left(\frac{T_w + 273}{T_{g\,iE} + 273}\right) + h_{1wW}[(T_w - T_{g\,iW}) - (T_a + 273)\ln\left(\frac{T_w+273}{T_{g\,iW}+273}\right)]\}A_b \quad (17)$$

The constant 0.933 represents the solar irradiation exchange constant, per hour water production for the proposed system have been calculated using the following equation.

$$M_w = \frac{q_{ew}}{L_v}3600 = \frac{h_{ew}(T_w - T_g)}{L_v}3600 \quad (18)$$

Latent heat of vaporization is expressed as [43]:

L_v=3.162510^6+[1-(7.61610^{-4} T_v)] For T_v> 70 °C

L_v=2.493510^6 [1-(9.477910^{-4} T_v)+1.313210^{-7} (T^2_v)-4.797410^{-3} (T^3_v)]

For T_v< 70 °C

Thermophysical properties of vapor, base fluid from references [31-34], and Nanoparticles (Table - 3) [37- 42].

Analysis based on matrices and cost conversion based on the life span of hybrid active double slope solar distiller unit using Al_2O_3 nanoparticles -

Energy matrices inform about the time of energy payback (EPT), energy payback factor (EPF), and life cycle conversion efficiency for the lifetime period (LCCE) [27].

Energy Payback Time (EPT) [27]

The time duration needed to recover total exhausted energy in manufacturing material can be expressed as

$$EPT_{(e)} = \frac{\text{Embodied Energy}(E_{in})}{\text{Annual energy output}(E_{out})} \quad (19)$$

$$EPT_{(ex)} = \frac{\text{Embodied Energy}(E_{in})}{\text{Annual exergy output}(E_{out})} \quad (20)$$

Energy Production Factor (EPF) [27]

It is reciprocal of EPT an ideal value of EPF on an annual basis to express the overall performance of solar still can be expressed as

$$EPF_{(e)} = \frac{\text{Overall Energy Output}(E_{out})}{\text{Embodied Energy}(E_{in})} \quad (21)$$

$$EPF_{(ex)} = \frac{\text{Overall Exergy Output}(E_{out})}{\text{Embodied Energy}(E_{in})} \quad (22)$$

Lifecycle Conversion Efficiency (LCCE) [27]

$$LCCE_{(e)} = \frac{E_{out} \times n - E_{in}}{E_{sol} \times n} \quad (23)$$

Where, E_{sol} represents annual solar energy (kWh) and n is the overall lifetime period.

Table 3. Thermophysical properties of Al2O3 nanoparticles [43]

Nanoparticles (NPs)	Density k_g/m^3	Thermal conductivity k_p (W/mK)	Specific heat C_p (j/k_g K)
Al_2O_3	$6.3*10^3$	17.6	550

Table 4. Daily, Monthly, and Annually yield of the proposed system

Month	whether condition type (a)			whether condition type (b)			whether condition type (c)			whether condition type (d)			Monthly yield
Jan	14.70	3	44.09	12.78	8	102.21	4.17	11	45.86	1.21	9	10.85	203.01
Feb	13.82	3	41.47	13.44	4	53.75	4.57	12	54.81	1.42	9	12.78	162.82
Mar	18.20	5	90.98	18.10	6	108.63	9.02	12	108.19	5.00	8	40.00	347.81
Apr	20.51	4	82.03	21.53	7	150.68	11.72	14	164.12	10.90	5	54.49	451.32
May	21.98	4	87.92	19.45	9	175.01	15.07	12	180.81	11.00	6	66.00	509.74
Jun	12.49	3	37.48	18.97	4	75.89	12.25	14	171.56	7.83	6	47.01	331.94
Jul	15.19	2	30.38	15.19	3	45.57	11.51	10	115.12	6.70	17	113.97	305.04
Aug	15.27	2	30.53	15.47	3	46.40	9.65	7	67.52	5.78	19	109.73	254.19
Sep	16.46	7	115.23	16.03	3	48.10	12.18	10	121.79	7.66	10	76.64	361.76
Oct	15.11	5	75.53	10.88	10	108.84	8.42	13	109.44	4.36	3	13.07	306.89
Nov	14.81	6	88.88	10.22	10	102.16	4.12	12	49.50	3.71	2	7.41	247.95
Dec	14.79	3	44.36	9.14	7	64.01	4.92	13	64.01	1.44	8	11.53	183.92
	Annual yield in (kg)												3666.39

Economic analysis is required to determine of hybrid active double slope solar still using Al_2O_3 nanoparticles-

It is economically feasible for System-A and System-B for both systems.

Capital cost

Tables 5 and 6 provide the system's fabrication cost.

System's Lifespan

It is considered for 15, 20, and 30 years.

Salvage value (S)

Salvage value (s) = 0.2 × Principal capital (PCC) (24)

PCC stands for principal capital cost

Cost of annual salvaging (ASC)

Annually salvage = S × shrinkage in a fund (SFF) (25)

SFF is a factor used for shrinkage

Yearly maintenance (AMC)

Yearly maintenance (AMC) =0.15×FAC (26)

FAC stands for the first annual cost.

Factor for capital recovery (CRF).

At a fixed rate of interest, it shows the present cost as a constant annual cost across time.

$$CRF = \frac{i(1+i)^n}{(1+i)^n - 1} \quad (27)$$

Shrinking fund factor

$$SFF = \frac{i}{(1+i)^n - 1} \quad (28)$$

Firstly annual cot estimated

FAC=PCC×CRF (29)

Net annual cost estimated

TAC= FAC+AMC-ASC (30)

Cost of distillate per kg obtained

$$Cost/kg = \frac{TAC}{\text{yield in life}} \quad (31)$$

Table 5. Embodied Energy of different components of system-A (N-PVT-DS-CPC-HE) and system-B (N-PVT-DS-FPC-HE)

Name of component	Embodied energy (kWh)	
	N-PVT-FPC-DS-HE (Previous system)[30]	N-PVT-CPC-DS-HE (Present system)
FRP body of DS	755.61	755.61
GI angle Iron	416.4	416.4
Glass cover	180.5	180.5
FPC (N=4)	2209.92	-
CPC (N=4)	-	3279.41
PV (glass-glass)	980	980
Copper heat exchanger	25.83	25.83
Nanoparticles (CuO)	17.82	17.82
Others	20	20
Total EE of system	4606.08	5675.57

Table 6. Capital investment and cost of different components of system-A (N-PVT-DS-CPC-HE) and system-B (N-PVT-DS-FPC-HE)

Parameters	N-PVT-FPC-DS-HE [32]		N-PVT-CPC-DS-HE	
	Cost	Previous system	Cost	Present system
	₹	$	₹	$
FRP body	10200	139.135	10200	139.135
Glass cover 2.05	1600	21.825	1600	21.825
Iron stand	1000	13.641	1000	13.641
Inlet /outlet nozzle	200	2.728	200	2.728
Iron clamp	250	3.410	250	3.410
Gaskets	200	2.728	200	2.728
Silicon gel	200	2.728	200	2.728
PVT-FPC (N=4) 8500	34000	463.784		0.000
PVT-CPC (N=4) 9000		0.000	36000	491.06
Motor and pump	1200	16.369	1200	16.369
Helically coiled heat exchanger	466.25	6.360	466.25	6.360
Fabrication and other cost	6000	81.844	6000	81.844
100 gmsAl_2O_3nanoparticles	7425	101.282	7425	101.282
Total cost of	62741.25	855.83	64741.25	883.11

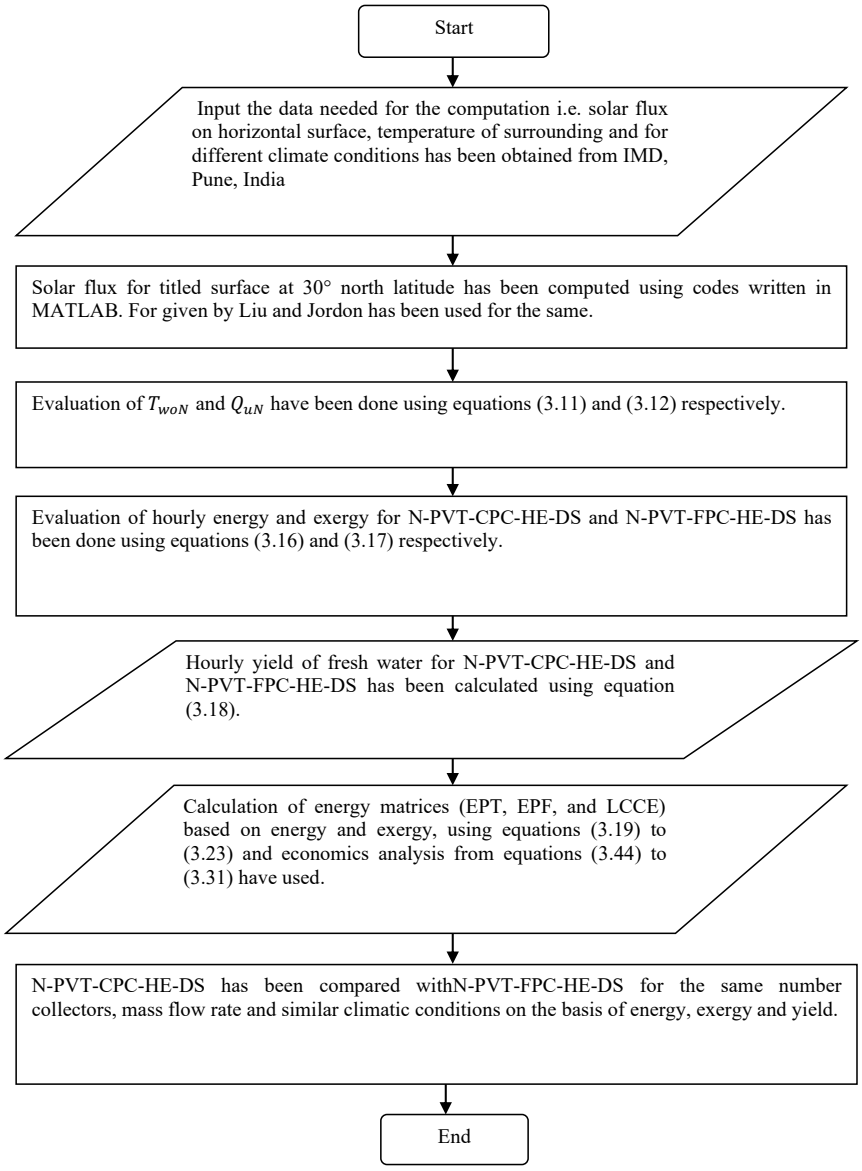


Fig. 2 Flow chart of the methodology adopted

Methodology to be Adopted

The following steps are included in the approach used to study the suggested system (Fig. 2):

Step-I

The proposed systems for the yearly are calculated using the Lui-Jordon formula for global and beam irradiation. Calculate daily solar radiation further by multiplying the number of days given in a month by the number of clear, hazy, hazy, cloudy, and cloudy days.

Step-II

The temperature of basin water is calculated based on hourly, monthly, and annual data, and all settings are tuned to maximize the collector's output temperature.

Step-III

Energy matrices such as efficiency of life cycle cost, Factor energy payback, energy payback time, and productivity have been evaluated.

Step-IV

Comparing proposed systems to the prior system using numerically computed values.

RESULT AND DISCUSSION

The solar irradiation on a flat area and surroundings temperature have been computed using

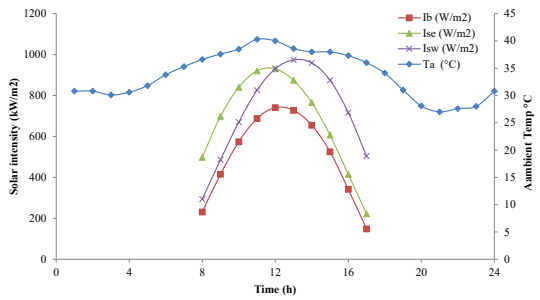


Fig. 3 Hourly variation in solar irradiation and ambient temperature for typical day of May

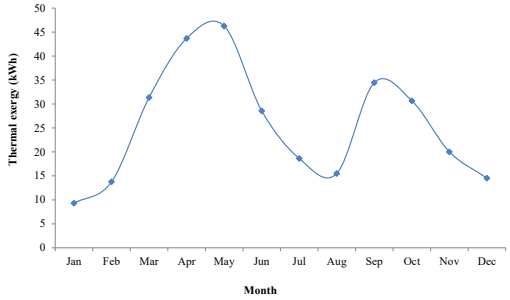


Fig. 4 Shows the thermal exergy of the proposed system-A

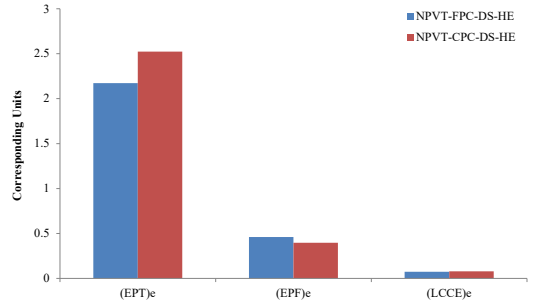


Fig. 5 Shows EPT, EPF, and LCCE for proposed and previous system for 15 years

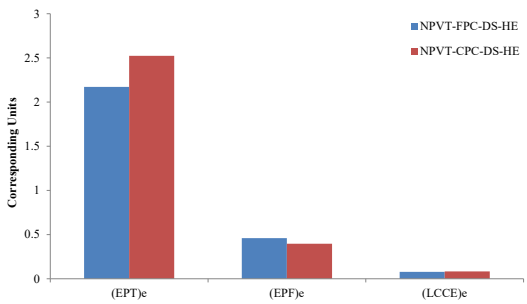


Fig. 6 Shows EPT, EPF, and LCCE for proposed and previous system for 20 years

IMD Pune data, India. By entering the pertinent data in MATLAB, the Liu and Jordan formula may be used to determine how much radiation was applied to N-PVT-CPCs. The values of beam radiation Ib, solar radiation Eastside ISE, Westside ISW, and ambient temperature Ta in kW/m^2 and ^{o}C, respectively, are shown in Fig. 3. Below Fig.4 shows a variation of the thermal exergy month-wise of proposed system-A.

Energy matrices and a life cycle cost conversion analysis are required for the hybrid active double slope solar distiller unit

Embodied energy (E_{in}), the conversion efficiency of the life cycle (LCCE), energy payback factor (EPF), and energy payback time for (N-PVT-DS-FPC-HE) and (N-PVT-DS-CPC-HE) are shown in Fig. 5, Fig. 6, and Fig. 7 for 15, 20 and 30 years respectively.

The number of energy matrices based on energy and exergy for proposed system-A is found that system-A is better to system-B. EPT based on energy and exergy is 16.16% and 17.84% higher, respectively. EPF based on energy and exergy is relatively less13.91% and 14.88%, respectively. LCCE based on energy and exergy is appreciably greater for system-A than system-B: 5.55%, 6.38%, and 7.15% for 15, 20, and 30 years of life considered, respectively.

Economic analysis is required to determine whether hybrid active solar still uses Nanoparticles (Al_2O_3)

It is economically feasible for system-A and System-B annually. Total yearly cost, fixed annual cost, yearly maintenance cost, and yearly water generation for system-A and system-B are shown during 30 years, at respective interest rates of 1%, 3%, and 5% respectively are represented in Fig. 8, and Fig. 9.

The cost of distillation relies on the rate of interest, as results over 30 years are shown in Fig.8. While the cost of distillate will reduce as system life increases, the price of distillate will climb annually as interest rates rise. For system-A and system-B, the cost of distillate is 0.69, 0.96, and 1.27 (/kg), and 0.72, 1.012, and 1.33 (/kg), respectively, over 30 years at interest rates of 1%, 3%, and 5%. According to the cost of distillation, it is concluded that the proposed system-A performs better than system B.

CONCLUSION AND FUTURE SCOPE

The proposed system and previous system performance have been studied using characteristic

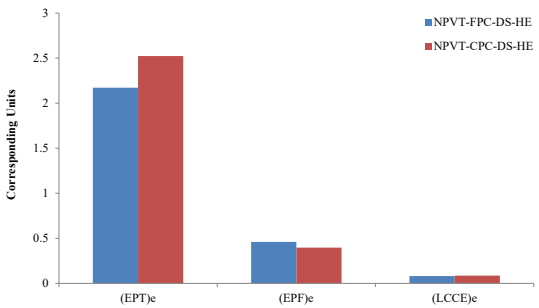


Fig. 7 Shows EPT, EPF, and LCCE for proposed and previous system for 30 years

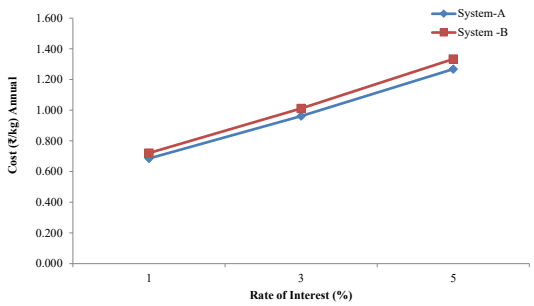


Fig. 8 Represents cost of distillate vs Interest rate

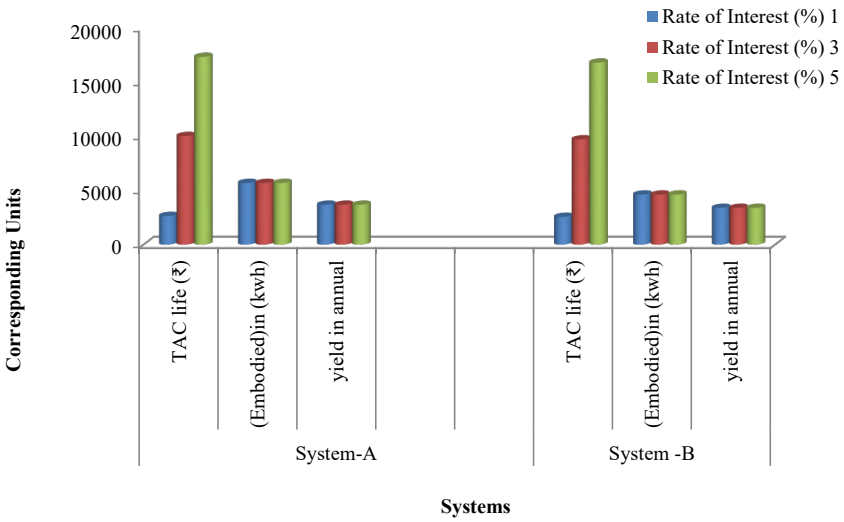


Fig. 9 Shows comparison for system-A and system-B based on TAC, Embodied Energy and Yield for 30 years

equations and Al_2O_3 nanoparticles and found better than the previous system.

Conclusions

The following conclusions are made by the annual analysis of the proposed systems with Al_2O_3 nanoparticles.

1. The proposed system-A gives better annual performance than system-B based on thermal exergy, energy, yield is 8.5%, and productivity 5.17% greater with Al_2O_3 nanoparticles.
2. Al_2O_3 nanoparticles-based system-A gives better results on annual performance and economics as compared to system-B.
3. Based on thermal energy, thermal exergy, energy, and exergy-based, Energy payback time (EPT) is 16.16%, Energy payback factor (EPF) is 13.91%, Efficiency of life cycle cost (LCCE) is 7.15% greater than the previous research, it is found that system-A outperforms system-B (previous).
4. System-A has a lower distillation cost than System B (previous). The yearly cost of distillate for systems A and B is 0.69, 0.96, and 1.27 (/kg), and 0.72, 1.012, and 1.33 (/kg), respectively, based on a 30-year basis at interest rates of 1%, 3%, and 5%. According to distillate cost, it is discovered that system-A performs better than system B.

Future scope

This work can be further expanded using research with PCM material in CPC up to a specific level.

1. Energy and exergy can be studied for different nanoparticles
2. energy matrices, EPT, EPF, and LCCE can be studied for different nanoparticles
3. Research on the environmental and economic benefits of various nanoparticles is possible. Different sizes, and shapes of nanoparticles can be investigated.
4. The partial covered 50%, 75%, and 100% can be studied.

Nomenclature

A_{ba} basin surface, in (m^2)
A_{ca} flat plate collector area under glazing, in (m^2)
A_{gE} eastside glass cover area, in (m^2)
A_{gW} westside glass cover area, in (m^2)
A_m PVT area, in (m^2)
C_p specific heat NPs, in (J/kgK)
C_{bf} base fluid specific heat, in (J/kgK)
C_{nf} specific heat nanofluid, in (J/kgK)
D_i FPC tube dia., in (m)
d_p Nps dia., in (nm)
F' the factor of collector efficiency
h_i coefficient of heat transfer glazing to absorbing plate, in ($W/m^2\,K$)
h_o heat transfer coefficient top of PVT to ambient air, in ($W/m^2\,K$)
h_{pw} heat transfer coefficient blackened plate to fluid, in ($W/m^2\,K$)
h_{bw} heat transfer coefficient liner to ambient, in ($W/m^2\,K$)
h_{ba} coefficient of heat transfer from basin to ambient, in ($W/m^2\,K$)
h_{CPC} coefficient of heat transfer convective in CPC, in ($W/m^2\,K$)
h_{HE} in heat exchanger convective heat transfer coefficient, in ($W/m^2\,K$)
h_{rwgE} eastside, water-to-glass heat transfer coefficient radiative, in ($W/m^2\,K$)
h_{rwgW} westside, water-to-glass heat transfer coefficient radiative, in ($W/m^2\,K$)
h_{cwgE} eastside, water-to-glass heat transfer coefficient radiative, convective in ($W/m^2\,K$)
h_{cwgW} westside, water-to-glass heat transfer coefficient radiative, convective in ($W/m^2\,K$)
h_{ewgE} eastside, water-to-glass heat transfer coefficient radiative, evaporative in ($W/m^2\,K$)
h_{ewgW} westside, water-to-glass heat transfer coefficient radiative, evaporative in ($W/m^2\,K$)
h_{1gE} eastside, coefficient of total heat transfer, in ($W/m^2\,K$)
h_{1gW} westside, coefficient of total heat transfer, in ($W/m^2\,K$)
h_{1wE} eastside water to glass cover coefficient of heat transfer, in ($W/m^2\,K$)
h_{1wW} westside water to glass cover coefficient of heat transfer, in ($W/m^2\,K$)
I_b on the collector solar irradiation, in (W/m^2)
I_{SE} eastside over-glass cover solar irradiation, in (W/m^2)
I_{SW} westside over glass cover solar irradiation, in (W/m^2)
K_g in glass heat conductivity($W/m\,K$)
K_p in absorbing plate heat conductivity ($W/m\,K$)
k_p in nanoparticles, heat conductivity($W/m\,K$)
K_{nf} in nanofluid heat conductivity ($W/m\,K$)
K_{bf} in base fluid heat conductivity($W/m\,K$)
L length of heat exchanger coil, in (m)
L_c glazing length collector, in (m)
L_i the thickness of insulation, in (m)
L_g cover of glass thickness, in (m)
L_m length of PVT, in (m)
L_p absorbing plate thickness, in (m)
M_w water mass, in (k_g)
M_{ew} yield, in (k_g)
m_f water flow rate, in (k_g/s)
PF_1 penalty factor 1
PF_2 penalty factor 2
PF_3 penalty factor 3
PF_c by glass cover for glazed portion, a penalty factor
P_{gi} partially saturated vapor pressure of glass cover, in (N/m^2)
Q_{uN} the heat transfer rate to N identical 25% PVT-CPC connected in series, in (kWh)
r_{11} outerdia of helical coiled heat exchanger tube, in (m)
r_{22} innerdia of helical coiled heat exchanger tube, in (m)
T_{giE} eastside, inside glass cover temperature, in ($°C$)
T_{giw} westside, inside glass cover temperature, in ($°C$)
T_{goE} eastside temperature of the outside glass cover, in ($°C$)

T_{goW} westside temperature of the outside glass cover, in $(°C)$
T_a surrounding temperature in(°C)
T_c the temperature of solar cell in(°C)
T_{cN} the average temperature of a solar cell, in (°C)
T_{fi} fluid inlet temperature in (°C)
T_{bf} in collector base fluid temperature, in (°C)
T_{foN} N^{th} collector outlet water temperature in (°C)
T_{nf} nanofluid temperature in(°C)
T_p absorbing plate temperature in (°C)
T_s sun temperature in(°C)
T_v vapor temperature in(°C)
T_w the temperature of basin water in(°C)
T_{wo} temperature of water at t=0, in(°C)
ΔT_{DSSS} temperature between nanofluid and base fluid, in (°C)
ΔT_{CPC} temperature between nanofluid to base fluid at PVT collectors outlet, in (°C)
ΔT_{HE} temperature between nanofluid to base fluid at the heat exchanger, in (°C)
ΔT temperature betweenT_w and $\frac{T_{giE}}{T_{giW}}$ for time(t), in (h)
U_{ba} sink line to surrounding, coefficient of overall heat transfer, in $(W/m^2\ K)$
U_{ga} condensing cover to surrounding, overall heat transfer coefficient, in $(W/m^2\ K)$
U_{gaE} condensing cover to surrounding, overall heat transfer coefficient eastside, in $(W/m^2\ K)$
U_{gaW} condensing cover to surrounding, overall heat transfer coefficient, in $(W/m^2\ K)$
U_{Lc} glazing to surrounding, overall heat transfer coefficient, in $(W/m^2\ K)$
U_{Lm} module to surrounding, overall heat transfer coefficient, in $(W/m^2\ K)$
U_{tca} solar cell to surrounding, overall heat transfer coefficient, in $(W/m^2\ K)$
U_{tpa} absorbing plate to surrounding, overall heat transfer coefficient, in $W/m^2\ K$
$U_{tc,p}$ solar cell to absorbing plate, overall heat transfer coefficient, in $(W/m^2\ K)$
X characteristic length, solar distiller unit, in (m)

Greek letters

α_g solar energy fraction absorbed by the condensing cover
α_b solar energy fraction absorbed by basin surface
α_f solar energy fraction absorbed by the fluid
α_c solar energy fraction absorbed by solar cell
β packing factor
β_p nanoparticles, coefficient of thermal expansion, in (K^{-1})
β_{nf} nanofluid coefficient of thermal expansion, in (K^{-1})
β_{bf} base fluid coefficient of thermal expansion, in (K^{-1})
$\emptyset_g$ nanoparticles fraction of volume, in (%)
μ_{bf} base fluid viscosity of dynamic, in (Ns/m^2)
μ_{nf} nanofluid viscosity of dynamic, in $(Ns/m^2))$
η_g collectors efficiency, in %
ρ_p nanoparticles density, in (Kg/m^3)
ρ_{nf} nanofluid density, in (Kg/m^3)
ρ_{bf} base fluids density, in (Kg/m^3)
τ_g transmitted fraction of glass cover

Subscripts

a ambient
a_n per annum
b basin
E eastside
e_n energy
e_x exergy
E_{in} embodied energy input
E_{out} embodied energy output
E_{sol} annual solar energy
f fluid
g_i condensing cover inside
g_o condensing cover outside
i interest rate
n life period
p particle
S_{ol} solar
th thermal
v vapor
W westside

Abbreviation

AMC maintenance costs annually
ASC salvage cost annually
BF base fluid
CRF capital recovery factor
C per liter yield price
CM carbon dioxide mitigation
DS double slope solar distiller unit
EPT payback time of energy
EPF payback factor of energy
FAC fixed annual cost
FPC collector, flat plate
HE heat exchanger
HTC coefficient of heat transfer
LCCE efficiency of life cycle conversion
NF nanofluid
NP nanoparticle
PCC primary capital cost
PVT photovoltaic thermal
R reflectors
SFF shrinking fund factor
S value of future salvage
TAC total annual cost
CPC compound parabolic concentrator
N-PVT-DS-CPC-HE, incorporating PVT-CPC double slope with N^{th} collector using a heat exchanger (helically coiled)

Appendix A

$$A_1 = C_1U_1 + C_2$$

$$C_1 = \alpha_g I_{SE} A_{gE} + U_{cgaE} A_{gE} T_a$$

$$C_2 = \alpha_g I_{SW} A_{gE} A_{gW} h_{EW} + U_{cgaW} A_{gE} T_a A_{gW} h_{EW}$$

$$U_1 = h_{1wE}\frac{A_b}{2} + h_{Ew}A_{gE} + U_{cgaE}A_{gE}$$

$$A_2 = (h_{1wE}U_2 + h_{1wW}h_{Ew}A_{gE})\frac{A_b}{2}$$

$$U_2 = U_{cgaW}A_{gW} + h_{1wW}\frac{A_b}{2} + h_{Ew}A_{gW}$$

$$B_1 = C_1'U_1 + C_2'$$

$$B_2 = (h_{1wW}U_1 + h_{1wE}h_{Ew}A_{gE})\frac{A_b}{2}$$

$$C_1' = \alpha_g I_{SW} A_{gE} + U_{cgaW} A_{gE} T_a A_{gW}$$

$$C_2' = \alpha_g I_{SE} A_{gE} A_{gW} h_{EW} + U_{cgaE} A_{gE} T_a A_{gW}$$

$$D_1 = m_f C_f\left(K_m^N - \frac{(1-e^{-z})}{(1-e^{-z}K_p^N)}\right)$$

$$D_2 = \left[\frac{(AF_R(\alpha\tau))1(1-K_p^N)}{(1-K_p)}\left(1 - \frac{e^{-z}}{(1-e^{-z}K_p^N)}\right)\right]$$

$$D_3 = \left[\frac{(AF_R(UL)1)(1-K_p^N)}{(1-K_p)}\left(1 - \frac{e^{-z}}{(1-e^{-z}K_p^N)}\right)\right]$$

$$E_1 = U_{gaE}[h_{EW} + h_{1bwW}(\frac{A_b}{2A_{gW}})$$

$$E_1' = U_{gaW}[h_{EW} + h_{1bwE}(\frac{A_b}{2A_{gE}})$$

$$E_2 = U_{gaW}(h_{EW} + U_{gaE})A_{gE}A_{gW}$$

$$E_2' = U_{gaE}(h_{EW} + U_{gaW})A_{gE}A_{gW}$$

$$H_1 = (U_{gaE} + U_{gaW})h_{1bwW}h_{1bwE}(\frac{A_b}{2})$$

$$H_2 = A_{gE}A_{gW}h_{EW}(U_{gaE}h_{1bwE} + U_{gaW}h_{1bwW})$$

$$H_3 = A_{gE}A_{gW}U_{gaE}U_{gaW}(h_{1bwE} + h_{1bwW})$$

$$H_4 = A_{gE}A_{gW}h_{EW}(U_{gaE}h_{1bwW} + U_{gaW}h_{1bwE})$$

$$H_{11}' = H_1 + H_2 + H_3 + H_4$$

$$H_{33}' = U_bA_b - D_1$$

$$H_{44}' == U_bA_b + D_3$$

$$K_{1E} = [h_{1bwW}\left(\frac{A_b}{2A_{gW}}\right) + h_{EW}\left(1 + \frac{1}{h_{1bwE}}\right) + U_{gaW}]A_{gW}\alpha_g A_{gE}h_{1bwE}$$

$$K_{1W} = [h_{1bwE}\left(\frac{A_b}{2A_{gE}}\right) + h_{EW}\left(1 + \frac{1}{h_{1bwW}}\right) + U_{gaE}]A_{gW}\alpha_g A_{gE}h_{1bwW}$$

$$K_{1E}' = [h_{1bwW}\left(\frac{A_b}{2}\right) + h_{EW}\left(1 + \frac{h_{1wE}}{h_{1wW}}\right)A_{gW} + U_{gaW}A_{gW}]\alpha_g A_{gE}h_{1wE}$$

$$K'_{1W} = [h_{1bwE}\left(\frac{A_b}{2}\right) + h_{EW}\left(1 + \frac{h_{1wE}}{h_{1wW}}\right)A_{gE} + U_{gaE}A_{gE}]\alpha_g A_{gW} h_{1wW}$$

$$P = U_1 U_2 - h_{EW}^2 A_{gE} A_{gW}$$

Expressions for K_k, $(A\,F_R(\alpha\tau))_1$ and $(A\,F_R U_L)_1$ used in Eq. (1) are as follows.

$$K_p = 1 - \frac{(AF_R(UL)1)}{m_f c_f}$$

$$K_m = 1 - \frac{(A_{rm}F_{Rm}(UL)m)}{m_f c_f}$$

$$U_{tca} = \left[\frac{1}{h_o} + \frac{L_g}{K_g}\right]^{-1} ; U_{tcp} = \left[\frac{1}{h_i} + \frac{L_g}{K_g}\right]^{-1} ;$$

$$h_o = 5.7 + 3.8V, \quad Wm^{-2}K^{-1} ; V = 1\ ms^{-1} ; h_i = 5.7, \quad Wm^{-2}K^{-1} ;$$

$$U_{tpa} = \left[\frac{1}{U_{tca}} + \frac{1}{U_{tcp}}\right]^{-1} + \left[\frac{1}{h'_i} + \frac{1}{h_{pf}} + \frac{L_i}{K_i}\right]^{-1} ;$$

$$h'_i = 2.8 + 3V', \quad Wm^{-2}K^{-1} ;$$

$$U_{L1} = \frac{U_{tcp}U_{tca}}{U_{tcp}+U_{tca}} ; U_{L2} = U_{L1} + U_{tpa} ; U_{Lm} = \frac{h_{pf}U_{L2}}{F'h_{pf}+U_{L2}} ; U_{Lc} = \frac{h_{pf}U_{tpa}}{F'h_{pf}+U_{tpa}} ;$$

$$PF_1 = \frac{U_{tcp}}{U_{tcp} + U_{tca}} ; PF_2 = \frac{h_{pf}}{F'h_{pf} + U_{L2}} ; PF_c = \frac{h_{pf}}{F'h_{pf} + U_{tpa}} ;$$

$$(\alpha\tau)_{1eff} = \rho(\alpha_c - \eta_c)\tau_g\beta_c\frac{A_{am}}{A_{rm}} ; (\alpha\tau)_{2eff} = \rho\alpha_p\tau_g^2(1-\beta_c)\frac{A_{am}}{A_{rm}} ;$$

$$(\alpha\tau)_{meff} = [(\alpha\tau)_{1eff} + PF_1(\alpha\tau)_{2eff}] ; (\alpha\tau)_{ceff} = PF_c.\rho\alpha_p\tau_g\frac{A_{ac}}{A_{rc}} ;$$

$$(AF_R(\alpha\tau))_1 = \left[A_cF_{Rc}(\alpha\tau)_{ceff} + PF_2(\alpha\tau)_{meff}A_mF_{Rm}(1 - \frac{A_cF_{Rc}U_{Lc}}{\dot{m}_f c_f})\right] ;$$

$$(AF_RU_L)_1 = \left[A_cF_{Rc}U_{Lc} + A_mF_{Rm}U_{Lm} + A_mF_{Rm}U_{Lm}(1 - \frac{A_cF_{Rc}U_{Lc}}{\dot{m}_f c_f})\right]$$

$$A_{rm} = b_rL_{rm} ; A_{am} = b_oL_{am} ;$$

$$A_cF_{Rc} = \frac{\dot{m}_f c_f}{U_{Lc}}\left[1 - \exp\left(\frac{-F'U_{Lc}A_c}{\dot{m}_f c_f}\right)\right] ;$$

$$A_mF_{Rm} = \frac{\dot{m}_f c_f}{U_{Lm}}\left[1 - \exp\left(\frac{-F'U_{Lm}A_m}{\dot{m}_f c_f}\right)\right] ;$$

$$z = \frac{2\pi r_{11}U_L}{\dot{m}_f c_f}$$

CONFLICT OF INTEREST

The authors hereby declare that there is no conflict of interest.

REFERENCES

[1] S. A. Lawrence, G. N. Tiwari, Theoretical evaluation of solar distillation under natural circulation with heat exchanger, Energy Convers. Manage., 30 (1990) 205-13. https://doi.org/10.1016/0196-8904(90)90001-F

[2] C. Popiel, and J. Wojtkowiak, Simple formulas for thermo-physicalpropertiesofliquidwaterforheattransfercalculations (from ° C to 150 °C). Heat Transfer Eng, 19 (1998) 87-101. https://doi.org/10.1080/01457639808939929

[3] B. C. Pak, Y. I. Cho., hydrodynamic and heat transfer study of dispersed fluids with submicron metallic oxide particles. Exp Heat Transfer: A J Therm. Energy Generation, Transport, Storage, Convers 11 (1998) 151-70. https://doi.org/10.1080/08916159808946559

[4] G. N. Tiwari, Solar energy: fundamentals, design, modelling and applications. New Delhi/NewYork: CRC Publication/ Narosa Publishing House; 2002

[5] K. S. Hwang, J. H. Lee, S. P. Jang , buoyancy-driven heat transfer of water-based Al2O3 nanofluids in a rectangular cavity, Int. J. Heat Mass Transfer, 50 (2007) 4003-10.

https://doi.org/10.1016/j.ijheatmasstransfer.2007.01.037
[6] O. O. Barden, Experimental study of the enhancement parameters on a single slope solar still productivity, Desalination, 209 (2007) 136-43. https://doi.org/10.1016/j.desal.2007.04.022
[7] G.N. Tiwari, A. K. Tiwari Solar distillation practice for water desalination systems, New Delhi, Anamaya Publishers, 2008.
[8] C. J. Ho, M. W. Chen, Z. W. Li, numerical simulation of natural convection of nanofluid in a square enclosure: effects due to uncertainties of viscosity and thermal conductivity, Int. J. HeatMass Transfer, 51 (2008) 4506-16. https://doi.org/10.1016/j.ijheatmasstransfer.2007.12.019
[9] T. P. Otanicar, J. Golden, Comparative environmental and economic analysis of conventional and nanofluid solar hot water technologies, Environ Sci. Technol., 43 (2009) 6082-7. https://doi.org/10.1021/es900031j
[10] H. E. Patel, T. Sundararajan, S. K. Das, An experimental investigation into the thermal conductivity enhancement in oxide and metallic nanofluids. J. Nanoparticle Res. 12 (2010) 1015-31. https://doi.org/10.1007/s11051-009-9658-2
[11] P. K. Singh, K. B. Anoop, T. Sundararajan, K. D. Sarit, entropy generation due to flow and heat transfer in nanofluids, Int. J. Heat Mass Transfer, 53 (2010) 4757-67. https://doi.org/10.1016/j.ijheatmasstransfer.2010.06.016
[12] M. G. J. D. Elzen, A. D Hof, A. M. Beltran, G. Grassi, M. Roelfsema, B. V. Ruijven, The Copenhagen accord: abatement costs and carbon prices resulting from the submissions, Environ. Sci. Policy, 14 (2011) 28-39. https://doi.org/10.1016/j.envsci.2010.10.010
[13] K. Khanafer and K. Vafai, A critical synthesis of thermo-physical characteristics of nanofluids, Int. J. Heat Mass Transfer, 54 (2011) 4410-28 https://doi.org/10.1016/j.ijheatmasstransfer.2011.04.048
[14] V. Khullar, H. Tyagi, A study on environmental impact of nanofluid based concentrating solar water heating system, Int J. Environ. Studies, 69 (2012) 220-32 https://doi.org/10.1080/00207233.2012.663227
[15] M. Faizal, R. Saidur, S. Mekhilef, M. A. Alim, Energy, economic and environmental analysis of metal oxides nanofluid for flat-plate solar collector, Energy Convers. Manage., 76 (2013)162-8. https://doi.org/10.1016/j.enconman.2013.07.038
[16] X. Liu, W. Chen, M. Gu, S. Shen, G. Cao, Thermal and economic analyses of solar desalination system with evacuated tubular collectors, Solar Energy, 93 (2013) 144-50. https://doi.org/10.1016/j.solener.2013.03.009
[17] A. E. Kabeel, Z. M. Omara, F. A. Essa, Enhancement of modified solar still integrated with external condenser using nanofluids, an experimental approach. Energy Convers.Manage., 78(2014) 493-8. https://doi.org/10.1016/j.enconman.2013.11.013
[18] T. Elango, A. Kannan, K. K. Murugavel, Performance study on single basin single slope solar still with different water nanofluids, Desalination, 360 (2015) 45-51. https://doi.org/10.1016/j.desal.2015.01.004
[19] Z. M. Omara, A. E. Kabeel, F. A. Essa, Effect of using nanofluids and providing vacuum on the yield of corrugated wick solar still. Energy Convers.Manage., 103 (2015) 965-72. https://doi.org/10.1016/j.enconman.2015.07.035
[20] G.N. Tiwari, J.K. Yadav, D.B. Singh, I.M. Al-Helal, A.M. Abdel-Ghaney, Exergoeconomic and enviroeconomic analyses of partially covered photovoltaic flat plate collector active solar distillation system, Desalination,367(2015) 186-96. https://doi.org/10.1016/j.desal.2015.04.010
[21] H. Sharon, K. S. Reddy, Performance investigation and enviro-economic analysis of active vertical solar distillation units, Energy, 84 (2015) 794-807. https://doi.org/10.1016/j.energy.2015.03.045
[22] L. Sahota, G. N. Tiwari, Effect of Al_2 O_3 NPs on the performance of passive double slope solar still, Solar Energy, 130 (2016) 260-72. https://doi.org/10.1016/j.solener.2016.02.018
[23] D. B. Singh, G. N. Tiwari, I. M. Al-Helal, V. K. Dwivedi, J. K. Yadav, Effect of energy matrices on life cycle cost analysis of passive solar stills, Solar Energy, 134 (2016) 9-22. https://doi.org/10.1016/j.solener.2016.04.039
[24] L. Sahota, G. N. Tiwari, Effect of nanofluids on the performance of passive double slope solar still: A comparative study using characteristic curve. Desalination, 388 (2016) 9-21. https://doi.org/10.1016/j.desal.2016.02.039
[25] D. B. Singh and G. N. Tiwari, Effect of energy matrices on life cycle cost analysis of partially covered photovoltaic compound parabolic concentrator collector active solar distillation system, Desalination, 397 (2016) 75-91. https://doi.org/10.1016/j.desal.2016.06.021
[26] S. W. Sharshir, G. Peng, L. Wu, N. Yang, F. A. Essa, A. H. Elsheikhd, Enhancing the solar still performance using nanofluids and glass cover cooling: Experimental study. Appllied Thermal Engg., 113 (2017) 684-93. https://doi.org/10.1016/j.applthermaleng.2016.11.085
[27] L. Sahota, G. N. Tiwari, Analytical characteristic equation of nanofluid loaded activedouble slope solar still coupled with helically coiled heat exchanger. Energy Convers. Manage., 135(2017) 308-26. https://doi.org/10.1016/j.ijleo.2017.01.063
[28] S.M. Saleha, A.M. Solimanb, M.A. Sharaf, V. Kaled, B. Gadgile, Influence of solvent in the synthesis of nano-structured Z_n O by hydrothermal method and their application in solar-still. J. Environ. Chem. Eng., 5 (2017) 1219-26. https://doi.org/10.1016/j.jece.2017.02.004
[29] W. Chen, C. Zou, X. Li, L. Li, Experimental investigation of S_i C nanofluids for solar distillation system: Stability, optical properties and thermal conductivity with saline water based fluid, International Journal of Heat Mass Transfer, 107 (2017) 264-70. https://doi.org/10.1016/j.ijheatmasstransfer.2016.11.048
[30] O. Mahian, A. Kianifar, S. Z. Heris, D. Wen, A. Z. Sahin, S. Wongwises, Nanofluids effects on the evaporation rate in a solar still equipped with a heat Exchanger. doi: http://dx.doi.org/10.1016/j.nanoen.2017.04.025.
[31] L. Sahota, Shyam, G. N. Tiwari, Energy matrices, enviroeconomic and exergoeconomic analysis of passive double slope solar still with water based nanofluids, Desalination, 409 (2017) 66-79. https://doi.org/10.1016/j.desal.2017.01.012
[32] D. B. Singh, G. N. Tiwari, Enhancement in energy metrics of double slope solar still by incorporating N identical PVT collectors, Solar Energy, 143 (2017) 142-161. https://doi.org/10.1016/j.solener.2016.12.039
[33] P. Joshi and G. N. Tiwari, Effect of cooling condensing cover on the performance of N identical photovoltaic thermal compound parabolic concentrator active solar still: a comparative study, International Journal of Energy and Environmental Engineering, 9 (2018) 473-498.

https://doi.org/10.1007/s40095-018-0276-6

[34] Dharamveer, Samsher, D. B. Singh, A. K. Singh, N. Kumar, Solar Distiller Unit Loaded withNanofluid-A Short Review. Lecture Notes in Mechanical Engineering, Springer, Singapore, (2019) 241-247, https://doi.org/10.1007/978-981-13-6577-5_24. https://doi.org/10.1007/978-981-13-6577-5_24

[35] S Kumar and D. Singh, Energy and exergy analysis of active solar stills using compound parabolic concentrator, International Research Journal of Engineering and Technology (IRJET), 6 (2019) 12.

[36] R. Shanker, D. Singh, D. B. Singh "Performance analysis of C.I. engine using biodiesel fuel by modifying injection timing and injection pressure" International Research Journal of Engineering and Technology (IRJET) 6 (2019) 12.

[37] A. K. Anup and D. Singh, FEA analysis of refrigerator compartment for optimizing thermal efficiency, International Journal of Mechanical and Production Engineering Research and Development, 10 (2020) 3, 3951-3972.

[38] S Kumar and D. Singh, Optimizing thermal behavior of compact heat exchanger, International Journal of Mechanical and Production Engineering Research and Development, 10 (2020) 3, 8113-8130.

[39] G.Zhang, N. D.Kaushika, S. C.Kaushik, R. K.Tomar, Advances in Energy and Built Environment, Springer Science and Business Media LLC, 2020. https://doi.org/10.1007/978-981-13-7557-6

[40] R. Dhivagar, M. Mohanraj, K. Hindouri, Y. Belyayev, Energy, exergy, economic and enviroeconomic (4E) analysis of gravel coarse aggregate sensible heat storage assisted single slope solar still, Journal of Thermal Analysis and Calorimetry, https://doi.org/10.1007/s10973-020-09766-w (2020). https://doi.org/10.1007/s10973-020-09766-w

[41] Dharamveer and Samsher, Comparative analyses energy matrices and enviro-economics for active and passive solar still, materialstoday: proceedings, https://doi.org/10.1016/j.matpr.2020.10.001 (2020). https://doi.org/10.1016/j.matpr.2020.10.001

[42] S. Arora, H. P. Singh, L. Sahota, M. K. Arora, R. Arya, S. Singh, A. Jain, A. Singh, Performance and cost analysis of photovoltaic thermal (PVT)-compound parabolic concentrator (CPC) collector integrated solar still using CNT-water based nanofluids, Desalination, 495 (2020) 114595. https://doi.org/10.1016/j.desal.2020.114595

[43] Dharamveer, Samsher, Anil Kumar, Analytical study of Nth identical photovoltaic thermal (PVT) compound parabolic concentrator (CPC) active double slope solar distiller with helical coiled heat exchanger using CuO Nanoparticles, Desalination and water treatment, 233 (2021) 30-51, https://doi.org/10.5004/dwt.2021.27526 https://doi.org/10.5004/dwt.2021.27526

[44] Dharamveer,Samsher, Anil Kumar, Performance analysis of N-identical PVT-CPC collectors an active single slope solar distiller with a helically coiled heat exchanger using CuO nanoparticles, Water supply, October 2021, SCI-E Index, IWA Publication. I.F 1.275,https://doi.org/10.2166/ws.2021.348 https://doi.org/10.2166/ws.2021.348

[45] M. Kumar and D. Singh, Comparative analysis of single phase microchannel for heat flow Experimental and using CFD, International Journal of Research in Engineering and Science (IJRES), 10 (2022) 03, 44-58.

[46] Subrit and D. Singh, performance and thermal analysis of coal and waste cotton oil liquid obtained by pyrolysis fuel in diesel engine, International Journal of Research in Engineering and Science (IJRES), 10 (2022) 04, 23-31.

[47} A. Shahsavar, P. Talebizadehsardari, M. Arıcı, (2022) Comparative energy, exergy, environmental, exergoeconomic, and enviroeconomic analysis of building integrated photovoltaic/thermal, earth-air heat exchanger, and hybrid systems. Journal of cleaner production. 362: 132510. https://doi.org/10.1016/j.jclepro.2022.132510 https://doi.org/10.1016/j.jclepro.2022.132510

J Mater Sci: Mater Electron (2023) 34:1662

Energy transfer induced colour tunable photoluminescence performance of thermally stable Sm^{3+}/Eu^{3+} co-doped Ba_3MoTiO_8 phosphors for white LED applications

Pooja Rohilla[1] and A. S. Rao[1,*]

[1] *Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110042, India*

Received: 13 March 2023
Accepted: 22 July 2023
Published online: 13 August 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

ABSTRACT

The combustion process has been successfully used to synthesize Ba_3MoTiO_8:-Sm^{3+} and Ba_3MoTiO_8:Sm^{3+}/Eu^{3+} co-doped phosphors for white light emitting diode (w-LED) applications. For the as-prepared phosphors, X-ray diffraction (XRD), scanning electron microscopy (SEM), fourier transform infrared spectroscopy (FT-IR), diffuse reflectance spectroscopy (DRS), and photoluminescence (PL) spectroscopy techniques have been used to investigate phase confirmation, morphological analysis, and PL properties. The PL emission spectra of singly Sm^{3+} doped and Sm^{3+}/Eu^{3+} co-doped phosphors recorded under 409 nm wavelength show intense emission in reddish–orange and red regions, respectively. For Sm^{3+}–Eu^{3+} co-doped phosphors, the intensity of Sm^{3+} peaks show a decrease with an increase in Eu^{3+} ion concentration, demonstrating the energy transfer from Sm^{3+} to Eu^{3+} ions. Quadrupole–quadrupole interaction is the mechanism responsible for energy transfer from Sm^{3+} ions to Eu^{3+} ions according to Dexter's hypothesis as applied to the PL spectra. This result is in consonance with the results of PL decay curves recorded under 409 nm excitation. The Commission Internationale de I' Eclairage coordinates calculated for singly Sm^{3+} doped and Sm^{3+}/Eu^{3+} co-doped phosphors show colour tunability from reddish–orange to pure red regions. All the results obtained finally revealed the utility of the titled phosphors as tunable red emitters needed to fabricate w-LEDs.

1 Introduction

In the current era, increasing human power consumption by humankind is taking a toll on natural resources. Most of the electricity produced is consumed by lighting devices and traditional lighting sources are being replaced by the white LEDs (w-LEDs), which are more efficient than the traditionally used incandescent and fluorescent lamps due to low power consumption, high brightness, compact

Address correspondence to E-mail: drsrallam@gmail.com

https://doi.org/10.1007/s10854-023-11025-1

design, and longer lifetime [1–4]. Researchers working in the field of phosphors are fiercely searching for novel and more efficient phosphor materials for the replacement of traditionally used materials in LEDs. Blue chips made of InGaN and yellow phosphor (YAG: Ce^{3+}) are used to create the commercially available white LEDs [5, 6]. But the light produced through this method has some shortcomings, such as a poor colour rendering index, a high correlated colour temperature, and mismatched warm white light due to the deficiency of the red component [7, 8]. The introduction of red/orange emitting phosphors can improve these limitations. So another approach that can be used for white light generation is either a combination of blue LED chips coated with green and red phosphors or UV LED chips coated with orange and cyan emitting phosphors. So, an orange-emitting phosphor in combination with a cyan-emitting phosphor under UV excitation can be a promising choice for researchers [9].

Recently, rare earth (RE) doped phosphor/glasses materials have been finding immense attention in the scientific community because of their vast applications in the fields of solid-state lighting technology, display devices, solar cells, etc. [10–13]. RE ion doping is usually preferred in phosphors due to their magnificent properties like high solubility, high luminescence efficacy, and large Stoke shifts [14–18]. Commercially, the red phosphor is fabricated with the sulfide-based Y_2O_2S: Eu^{3+} material, which is excited in the near-UV range [19]. The material Y_2O_2S: Eu^{3+} is chemically unstable, decreasing its luminous efficiency [20]. Therefore, as a replacement for this, the development of a highly efficient red phosphor is required. The titanate-based phosphors have some distinct properties, like a high refractive index and a wide bandgap, because of which they have good emission properties and therefore can be preferred as a stable host compound among the other phosphor materials. In addition to this, low synthesis temperature, physical and thermal stability and the broad excitation band in the UV region of molybdenum (Mo) present in the host can also improve the luminous properties of the BMT phosphor [21–23]. Also, the chromaticity of the titanate phosphors can be varied with the doping elements in 3d and 4f systems. Recently, co-doping of RE's in host lattices has driven interest due to their attractive energy transfer mechanism and flexible colour tuning properties. Among all the RE ions, the Eu^{3+} and Sm^{3+} ions show promising results and give emission in the reddish–orange and pure red regions, respectively [20–23]. In a broad emission range that spans the UV to visible regions, they exhibit good electrical and optical properties. Here, optical transitions occur between a variety of f states, and the filled 5s and 5p subshells effectively shield the f electrons from their surroundings [11, 24–26]. The Eu^{3+} is an effective activator ion due to its high colour purity and its intense red emission (615 nm) corresponding to transition $^5D_0 \rightarrow {}^7F_2$ [24]. The Sm^{3+} ion can act as a good sensitizer because of the lower energy difference between Sm^{3+} ($^4G_{5/2}$) and Eu^{3+} ($^5D_0$) and can transfer its excitation energy to Eu^{3+}. So, co-doping of Sm^{3+}/Eu^{3+} ions in a host lattice can broaden the excitation bands and helps in enhancing the luminescence efficiency for the red color emission in w-LEDs. Sm^{3+} ions can also act as a good activator when doped singly in some host compounds and gives reddish-orange emission. Such singly Sm^{3+} ions doped phosphors are used in such devices where the emission corresponding to the reddish-orange regions is needed. Therefore, in this article, we have reported our work on singly Sm^{3+}doped and Sm^{3+}/Eu^{3+} co-doped barium molybdenum titanate (BMT) based host lattice Ba_3MoTiO_8.

XRD, SEM, and Energy Dispersive X-Ray Analysis (EDX) analysis have all been used to get the information relevant to structure and morphology. FT-IR studies were used to understand various functional groups present in the titled BMT phosphor. Diffuse Reflectance Spectral (DRS) studies were conducted to understand the band gap of Sm^{3+} and Sm^{3+}/Eu^{3+} co-doped BMT phosphors. The PL emission studies were conducted on Sm^{3+} and Sm^{3+}/Eu^{3+} co-doped BMT phosphors to understand the emission properties and color tunability possessed by the titled phosphor. Further PL decay studies were conducted to understand the energy transfer mechanism responsible for the enhancement of PL intensity and color tunability. The PL emission spectral information of Sm^{3+} and Sm^{3+}/Eu^{3+} co-doped BMT phosphors was further used to calculate the CIE chromaticity coordinates to understand the utility of the titled phosphors for w-LED applications.

2 Experimental method and characterization techniques used

The preparation of the BMT phosphor was done through the combustion method. In our previous work, we have reported the superiority of the combustion method used to prepare BMT phosphor because of relatively high luminescence efficiency, crystallinity, and better morphology [25]. This has prompted us to use the combustion method to prepare BMT phosphor co-doped with samarium and europium ions to study color tunability. The precursors like ammonium molybdate (CDH, 98%) and barium nitrate (Fisher Scientific, 99%) were dissolved in DI water in the stoichiometric ratio, individually as per the equation.

$$(105 - x - y)\ Ba(NO_3)_2 + 35\ C_{10}H_{14}O_5Ti + 5\ (NH_4)_6Mo_7O_{24} \times 4H_2O + 48\ C_2H_5NO_2 + 131\ H_2O + xSm(NO_3)_3 + yEu_2O_3 = 35\ Ba_3MoTiO_8 + 446\ CO_2 + 288\ NH_4.$$

At the same time, titanium (IV) oxyacetylacetonate (Sigma-Aldrich, 90%) was dissolved in methanol at 55 °C to obtain a clear solution. Subsequently, solutions for doping ions were prepared by adding Eu_2O_3 in HNO_3 (Sigma-Aldrich, 99.9%) at 75 °C and $Sm(NO_3)_3$ (Sigma-Aldrich, 99.9%) in DI water at room temperature, respectively. After this, all the above solutions were mixed together at 100 °C for 1 h. After homogeneous mixing, stoichiometric amount of glycine (Fisher Scientific, AR Grade) was added, and the mixture was stirred up continuously for 2 h. The obtained mixture was placed on a preheated hot plate at 500 °C for the dehydration process. As a result, the grey foamy powder was obtained after the release of a large number of gases. Finally, the foamy powder was crushed into the fine powder in an agate mortar, and the homogeneous powder was sintered at 1100 °C for 10 h to obtain a pure phase under normal conditions. After the sintering process, powder was grounded again for 15 min to make it fine and homogeneous. Figure 1 depicts the detailed steps involved in the synthesis process.

The phase identification and confirmation of all the as-synthesized BMT phosphors was determined by an X-ray diffractometer (Bruker D8 Advanced X-Ray Diffractometer) with Cu-Kα radiation (λ = 1.54 Å) in the range of 2θ = 20°–70°. To analyze the structural parameters, Rietveld refinement was done using FullProf Suite software. The surface morphology and elemental distribution were checked by SEM on a JEOL 7610F Plus operated at 5 kV. To determine the various vibrational functions and groups present in the system, FT-IR was performed by Perkin Elmer spectrum 2. The diffuse reflectance spectral (DRS) studies were conducted on a Jasco-made V-770 Spectrophotometer at room temperature. A Jasco FP-8300 spectrofluorometer was used to record the PL excitation and emission spectra with a Xenon lamp serving as the excitation source. On a Hameg Instruments HM 1507 digital oscilloscope operating at 150 MHz, PL decay curves were captured. The ocean optics setup was used to record the temperature-dependent PL experiments with a resolution of 0.1 nm under a Xenon lamp excitation source.

3 Results and discussion

3.1 Structural and morphological analysis

Figure 2(a) and (b) show the XRD patterns of un-doped, single-doped BMT: xSm^{3+} (x = 0.5, 1, 1.5, 2, and 2.5 mol%) and co-doped BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1, 2, 3, 4, 5, 6 mol%) phosphors along with standard JCPDS data (JCPDS No. 38-0254). From Fig. 2, it can be seen that all the diffraction patterns were in good agreement with the standard JCPDS patterns with rhombohedral crystal structure and space group R-3 m (166) without the formation of any secondary crystallite phase. This shows the successful incorporation of dopant ions into the host lattice and single-phase confirmation of all the phosphors. By zooming in on the main intense peak of the XRD data, it is found that there is some shifting in the peak position towards the higher angle side with the introduction of Sm^{3+} and Eu^{3+} ions in the host lattice when compared with the un-doped sample. This shifting may be caused by the difference in the ionic radii of the Ba^{2+} ions (ionic radii = 1.49 Å) in comparison to the dopant ions Sm^{3+} (ionic radii = 1.09 Å) and Eu^{3+}(ionic radii = 1.08 Å). Figure 3(a) and (b) shows the Rietveld refinement fitting for the xSm^{3+} (x = 1.5 mol%) and xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1 mol%) doped BMT phosphors using Full Proof suite software. In our earlier work, we performed Rietveld refinement on an un-doped BMT sample [23]. All the structural parameters extracted from the refinement fitting were listed in Table 1. The goodness of the fit (χ^2) value is less than

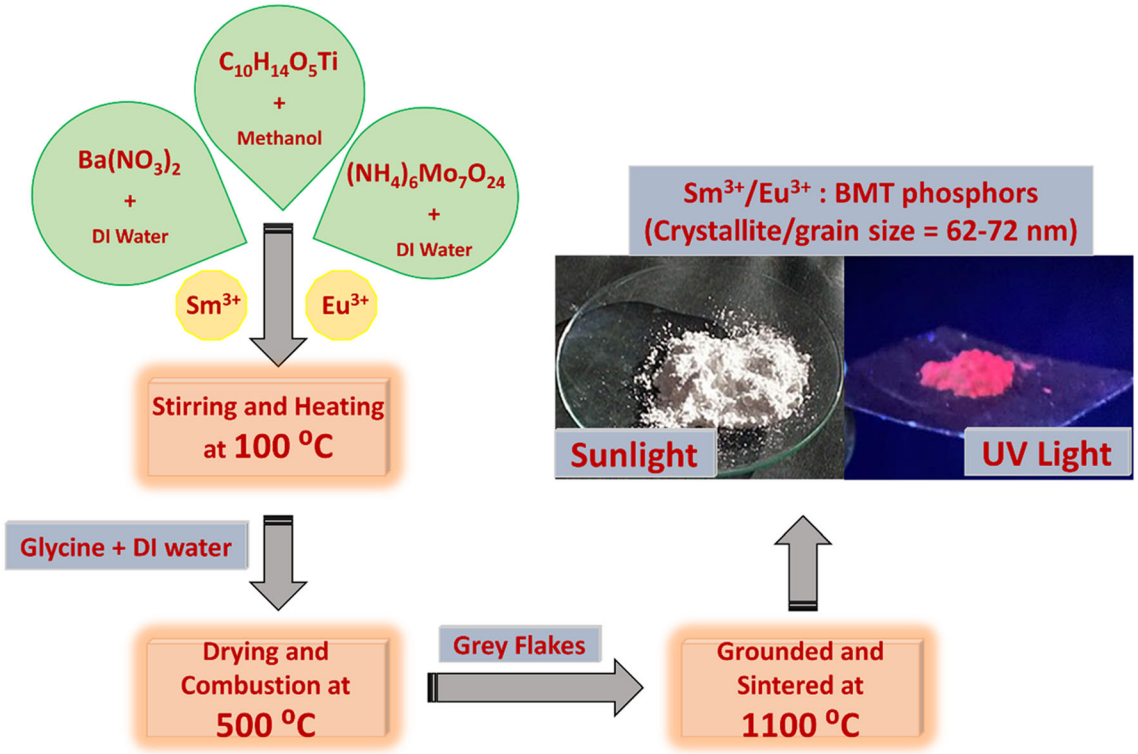


Fig. 1 Flow chart depicting synthesis steps involved in the combustion method

1, indicating no significant deviation in lattice parameters and the successful formation of single phase. The average crystallite size of the single and co-doped phosphors was calculated by the Debye–Scherrer formula given by:

$$D = \frac{s\lambda}{\beta_T \cos\theta},$$

where s, λ, β_T, and θ are shaping factor constant, the wavelength of X-rays, full-width at half maxima (FWHM), and diffraction angle of XRD peaks, respectively. Here D is the crystallite size. The crystallite size corresponding to the most intense peak calculated for single Sm^{3+} ion-doped phosphors lies in the 60–70 nm range, and for Sm^{3+}/Eu^{3+} co-doped phosphors, it lies in the range of 62–72 nm. As we know, the broadening of the XRD peak is due to the crystallite size in addition to the strain present in it. To calculate the lattice strain present in the crystal, a Williamson–Hall plot has been used using the equation [26]

$$\beta_T \mathrm{Cos}\theta = \frac{K\lambda}{D} + \varepsilon(4\mathrm{Sin}\theta).$$

The symbol ε represents the strain present in the lattice, and the other symbols have the exact same meaning as discussed above. A linear graph has been plotted between 4Sinθ and β_T Cosθ by considering the major peaks in the XRD data, as shown in Fig. 3(c) and (d). The slope of the fitted line gives the value of strain, and from the intercept value, we calculate the crystallite size. The average crystallite size for singly Sm^{3+} doped and Sm^{3+}/Eu^{3+} co-doped phosphors was found to be 98 nm and 88 nm, respectively. The value of strain was found in the range of 0.000170–0.000353 for Sm^{3+} doped phosphors, and for Sm^{3+}/Eu^{3+} co-doped phosphors, it lies in the range of 0.000267–0.0007.

Additionally, SEM and EDX have been performed to further investigate the microscopic morphology and elemental composition of the prepared phosphors. Figure 4(a) and (b) represent the SEM micrographs of single doped BMT: xSm^{3+}(x = 1.5 mol%) and co-doped BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1 mol%) phosphors. The particles of these phosphors are irregular in shape with much agglomeration. These irregularities in shape and size may be due to the unequal distribution of mass and temperature during the sintering as well as due to the

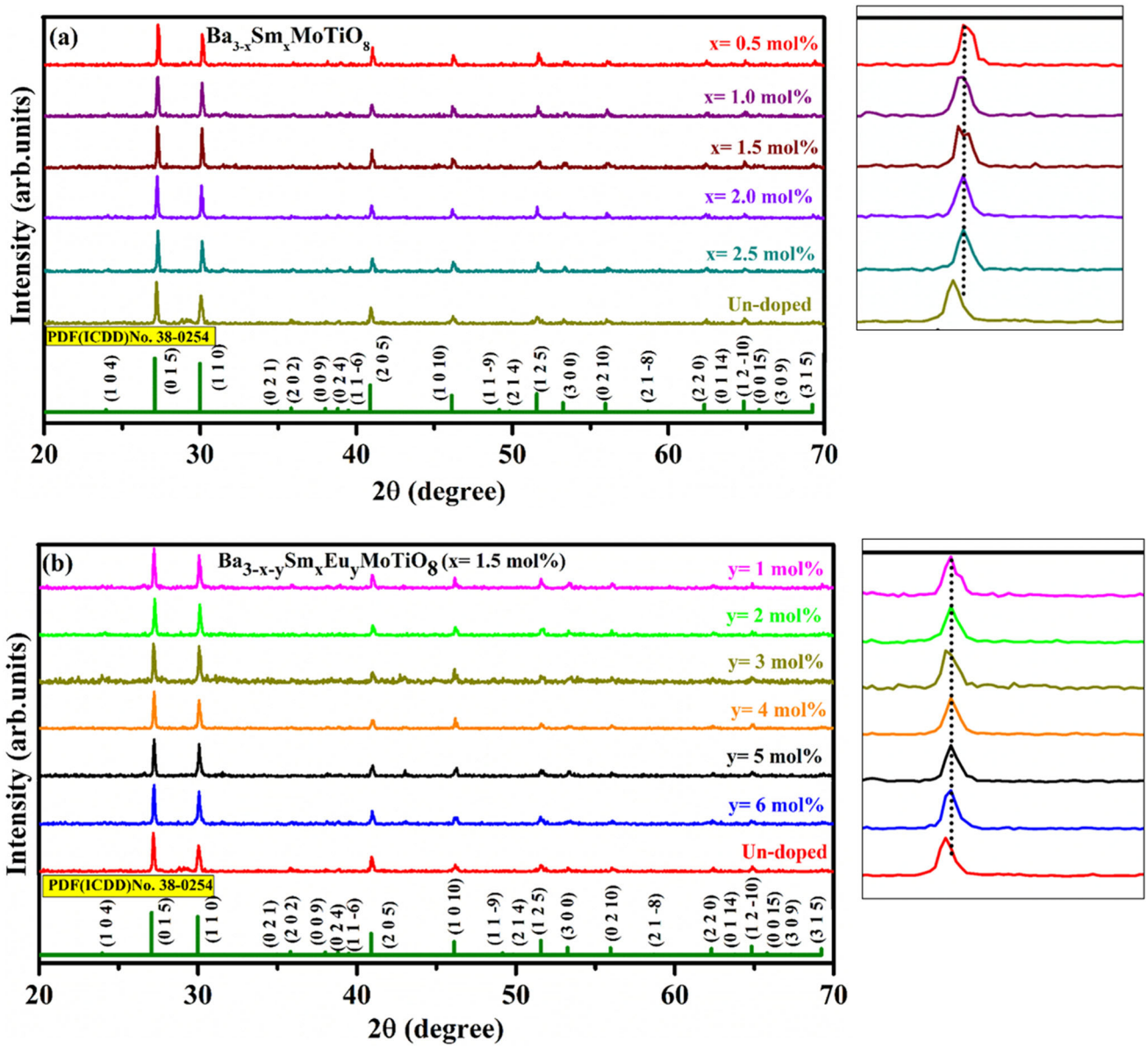


Fig. 2 (**a**), (**b**) XRD patterns of an un-doped, single doped BMT: xSm^{3+}(x = 0.5, 1.0, 1.5, 2.0 and 2.5 mol%) and co-doped BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol% and y = 1, 2, 3, 4, 5 and 6 mol%) phosphors with standard JCPDS

manual crushing of phosphor after sintering [27]. The size of particles for single-doped and co-doped phosphors are in the range of 0.5–5.0 μm and 0.1–6.0 μm respectively. Figure 4(c) and (d) demonstrate the corresponding EDAX images of these phosphors. It can be seen from both the EDX images that all elements are present in the phosphor lattice that have been used during synthesis.

Figure 5 shows the elemental mapping of co-doped BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1 mol%) phosphor. From the elemental mapping images, it was clear that barium, molybdenum, titanium, samarium, and europium elements were uniformly distributed in the particles, which shows the successful synthesis of phosphor.

3.2 FT-IR analysis

FT-IR spectra of un-doped, singly doped BMT: xSm^{3+} (x = 1.5 mol%) and co-doped BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1 mol%) phosphors were recorded in attenuated total reflectance mode (ATR), as shown in Fig. 6. The spectra were in the range 400–4000 cm^{-1} to analyse the different functional groups that are

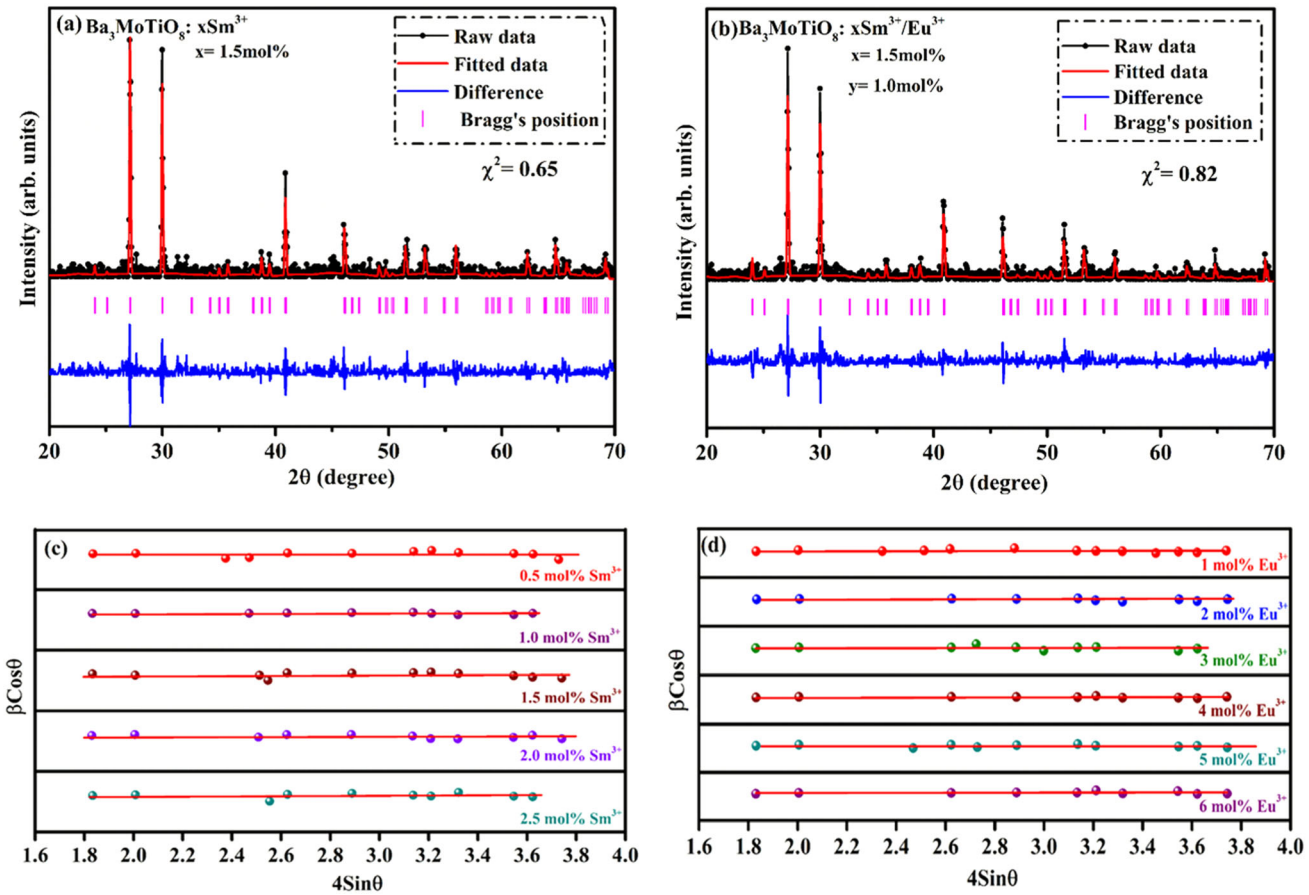


Fig. 3 (**a**), (**b**) Rietveld refinement for BMT: xSm^{3+}(x = 1.5 mol%) and BMT: xSm^{3+}/Eu^{3+} (x = 1.5 mol%, y = 1.0 mol%) co-doped phosphors respectively and (**c**), (**d**) correspondong W–H plots for all concentrations of Sm^{3+} and Eu^{3+} ions

Table 1 Calculated structural lattice parameters through Rietveld refinement

Parameters	Standard data JCPDS (038-0254)	BMT: xSm^{3+} x = 1.5 mol%	BMT: xSm^{3+}/yEu^{3+} x = 1.5 mol% y = 1 mol%	BMT (un-doped)
a (Å)	5.9570	5.9596	5.9568	Ref. [23]
b (Å)	5.9570	5.9596	5.9568	
c (Å)	21.2900	21.2903	21.2936	
α = β	90	90	90	
γ	120	120	120	
χ^2	–	0.65	0.82	
Cell volume(Å^3)	654.28	654.84	654.32	

present along with their stretching and bending vibrations. The infrared bands of all the phosphors occupy an almost similar position with or without doping. The absorption band at 530 cm^{-1} belongs to the stretching vibrations of Ti–O groups present in the system [22]. The bands at 678 and 818 cm^{-1} are due to the MoO_4^{2-} tetrahedrons ascribed to their Mo–O and Mo–O–Mo stretching vibrations [28, 29]. The bands at 1068 and 1421 cm^{-1} demonstrate C–N stretching and C=C stretching, respectively [30, 31]. The band at 1230 cm^{-1} describes the N–H bending absorption [32]. The two bands at 2897 and 2996 cm^{-1}

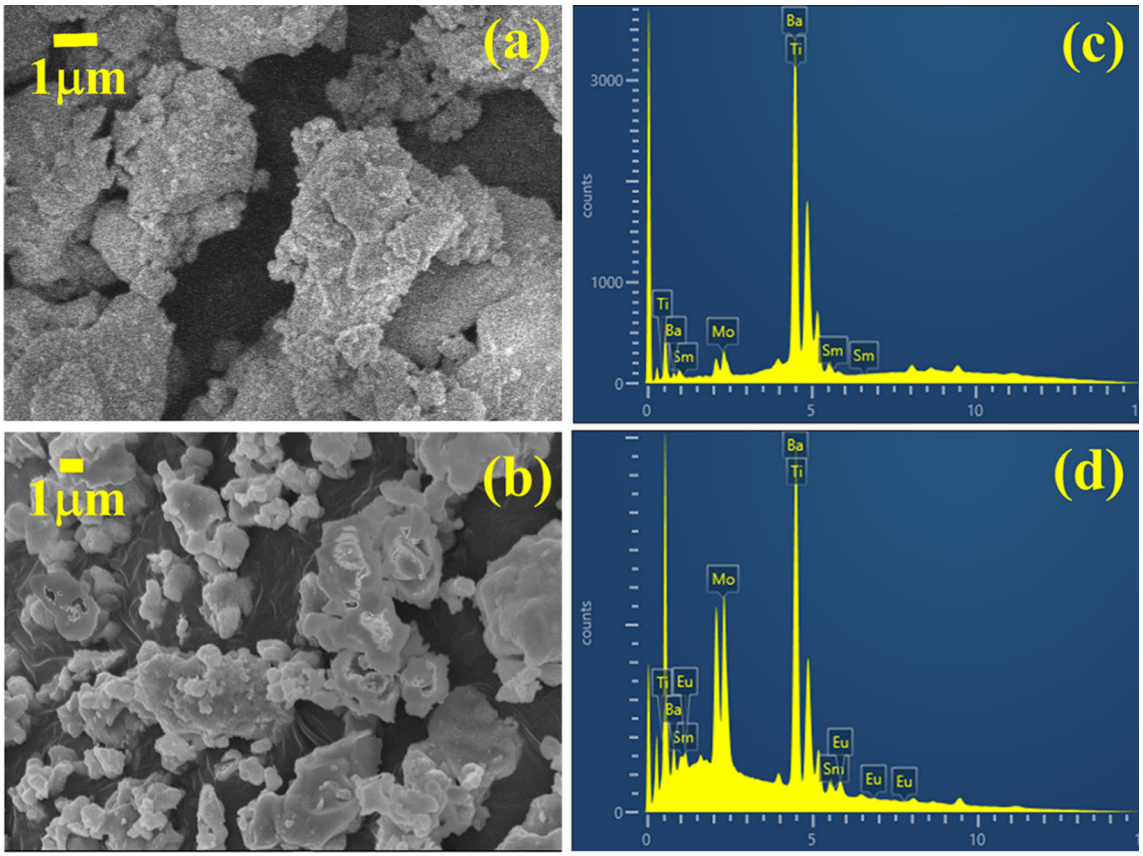


Fig. 4 Scanning electron microscope images (**a**), (**b**) of BMT: xSm^{3+} doped (x = 1.5 mol%) and BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol% and y = 1 mol%) co-doped phosphors and (**c**), (**d**) corresponding EDX images

represent C–H bonding [33]. The shortest absorption band at 3670 cm^{-1} can be assigned to the stretching mode of O–H groups in absorbed water molecules [34].

3.3 DRS spectra

Diffuse reflectance spectra were recorded for the BMT: xSm^{3+} (x = 0.5, 1, 1.5, 2 and 2.5 mol%) and co-doped BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1, 2, 3, 4, 5, 6 mol%) phosphors. To find the optical band gap, the Kulbelka–Munk function was applied to the DRS data given by equation [35]

$$F(R) = \frac{(1-R)^2}{2R} = \frac{A}{S}.$$

The Kulbelka–Munk function, reflectance coefficient, molar absorption coefficient, and scattering coefficient, respectively, are represented in this equation as F(R), R, A, and S. The relationship between the energy band gap (Eg) value and the absorption coefficient (α) was later determined using the Tau equation.

$$\alpha h\upsilon = C(h\upsilon - \mathrm{Eg})^n.$$

In this equation, n is a transition dependent constant with values of 1/2, 2, 1/3, and 3 for direct allowed, indirect allowed, direct forbidden, and indirect forbidden, respectively. Further, h stands for Planck's constant, C for an energy-independent constant and υ for the frequency of light.

Since F(R) is directly proportional to α, the above tauc equation becomes [36]

$$F(R)h\upsilon = C(h\upsilon - \mathrm{Eg})^n.$$

Now a graph was plotted between $[F(R)h\upsilon]^2$ vs. energy hυ for the n = 1/2 transition, as shown in Fig. 7. By extrapolating from the linear section of the figure, the intercept on the x-axis provides an approximation of the band gap value. The approximate band gap values for singly Sm^{3+} doped BMT phosphors lie in the range 3.3–3.5 eV, and for co-

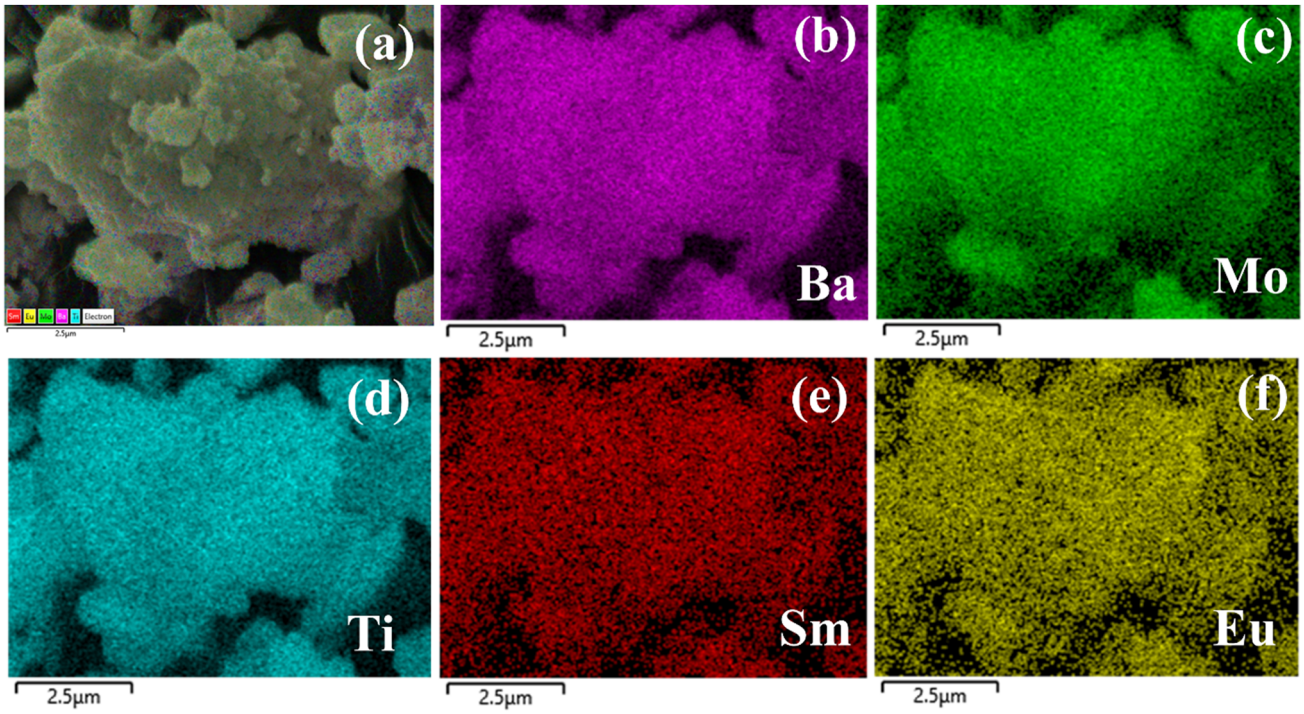


Fig. 5 Elemental mapping of BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1 mol%) co-doped BMT phosphor (**a**) mixed mapping (**b**–**f**) individual of Barium (Ba), Titanium (Ti), Molybdenum (Mo), Samarium (Sm) and Europium (Eu) elements

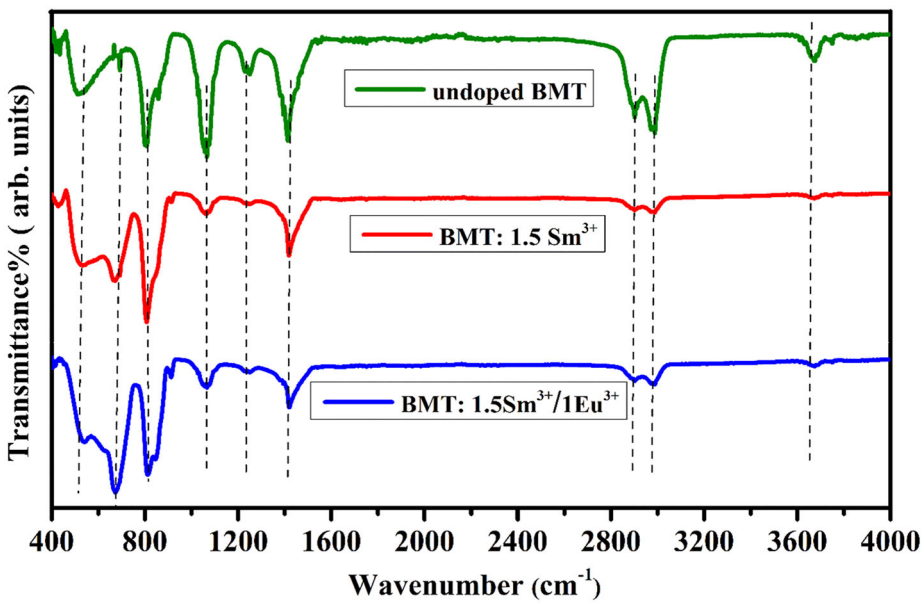


Fig. 6 FT-IR spectra of an undoped, xSm^{3+} doped(x = 1.5 mol%), and co-doped xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1 mol%) phosphors

doped BMT: Sm^{3+}/Eu^{3+} phosphors lie in the range 3.0–3.25 eV, respectively [37]. The band gap energy decreases with the addition of co-doping of Eu^{3+} ions. This results in more number of atoms getting excited to the upper states and thereby resulting a stronger emission when they are falling down to the lower energy states. The band gap values obtained in the present work are in consonance with the values reported for other phosphors doped with Sm^{3+} and Eu^{3+} ions [38, 39].

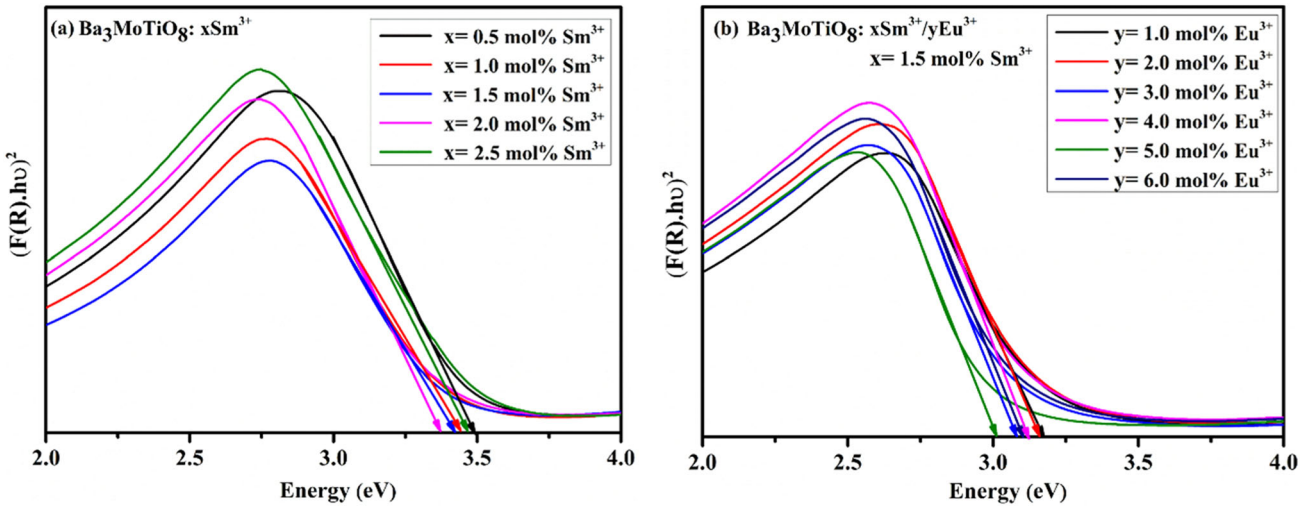


Fig. 7 $[F(R)hv]^2$ versus hν plot for optical band gap calculation of (**a**) BMT: xSm^{3+}(x = 0.5, 1.0, 1.5, 2.0 and 2.5 mol%) and (**b**) BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol% and y = 1, 2, 3, 4, 5 and 6 mol%) phosphors

3.4 PL studies

3.4.1 *PL studies of BMT: xSm^{3+}(x = 1.5 mol%) phosphor*

Figure 8(a) and (b) show the excitation spectrum and emission spectra of BMT: xSm^{3+}(x = 1.5 mol%) phosphor, respectively. A 596 nm emission wavelength was used to record the excitation spectra, which ranged from 325 to 500 nm. The spectrum comprises several peaks at 364, 381, 409, 421,442 and 465 nm corresponding to the transitions $^6H_{5/2} \rightarrow {}^4D_{3/2}$, $^6H_{5/2} \rightarrow {}^4D_{1/2}$, $^6H_{5/2} \rightarrow {}^4F_{7/2}$, $^6H_{5/2} \rightarrow {}^4M_{19/2}$, $^6H_{5/2} \rightarrow {}^4G_{9/2}$ and $^6H_{5/2} \rightarrow {}^4I_{11/2} + {}^4I_{13/2} + {}^4M_{15/2}$ respectively [21, 40–42]. Out of these peaks, the most intense peak in the excitation spectrum is at 409 nm, so we recorded the emission spectra for BMT: xSm^{3+}(x = 0, 1, 1.5, 2 & 2.5 mol%) series at this excitation wavelength. The emission spectra of this series recorded in the range of 540–700 nm comprise three bands at wavelengths 562, 596, and 643 nm corresponding to a transition from the excited state $^4G_{5/2}$ of Sm^{3+}ions to ground states $^6H_{5/2,}{}^6H_{7/2}$ and $^6H_{9/2}$ respectively [43, 44]. The intensity of emission peaks increases with Sm^{3+}ions concentration up to 1.5 mol%; beyond that, emission intensity decreases, which is attributed to the phenomenon of concentration quenching. Concentration quenching is the phenomenon in which if we increase the activator ions higher than a certain value, the number of non-radiative transitions rises through cross-relaxation processes and results in low PL intensity. In the present work, it was observed that, 1.5 mol% of Sm^{3+} ions produce concentration quenching as shown in Fig. 8(b). The most intense band at 596 nm wavelength corresponds to reddish–orange emission under UV excitation. So Sm^{3+} doped BMT phosphors can be used with cyan-emitting phosphors by coating them with UV LED as reddish–orange colour emitting devices, and by co-doping them with Eu^{3+} ions, they can be used to obtain pure red emission.

3.4.2 *PL studies of single doped Eu^{3+} and Sm^{3+}/Eu^{3+} co-doped BMT phosphors*

The PL excitation and PL emission spectra of BMT: yEu^{3+}(y = 1 mol%) phosphor were reported in Fig. 8(c) at λ_{em} = 615 nm and λ_{ex} = 395 nm, respectively. The excitation spectrum recorded in the range 300–550 nm comprises peaks at wavelengths 306, 360, 381, 395, 415, and 465 nm corresponding to the charge transfer band (CTB), $^7F_0 \rightarrow {}^5D_{4,}{}^7F_0 \rightarrow {}^5L_{7,}{}^7F_0 \rightarrow {}^5L_{6,}{}^7F_0 \rightarrow {}^5D_{3,}$ and $^7F_0 \rightarrow {}^5D_2$ transitions of Eu^{3+} ions respectively [45]. The emission spectrum was recorded using the most intense peak of the excitation spectrum (λ_{ex} = 395 nm) in the range of 550–750 nm. The emission spectrum consists of five peaks at wavelengths 579, 596, 615, 653, and 701 nm corresponding to the transition of Eu^{3+} ions from excited state $^5D_0$ to $^7F_{0,}{}^7F_{1,}{}^7F_{2,}{}^7F_{3,}$ and $^7F_4$ states, respectively [46]. The transition $^5D_0 \rightarrow {}^7F_1$ is a magnetic dipole (MD) in nature, whereas $^5D_0 \rightarrow {}^7F_2$ is an electric dipole (ED). In the emission graph, if the ED transition is ruling over the MD transition, that means the

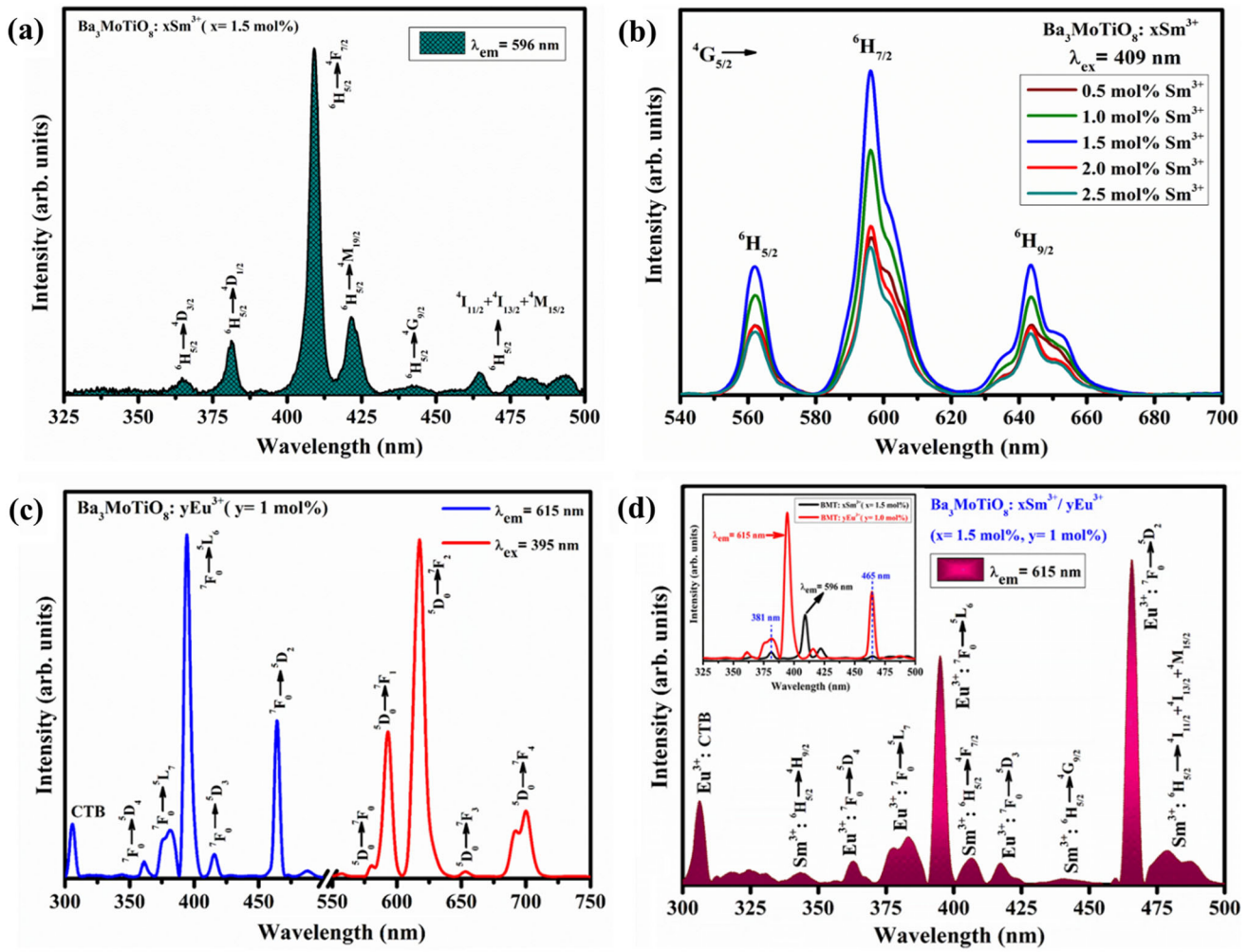


Fig. 8 (**a**) PL excitation spectrum of BMT: xSm^{3+}(x = 1.5 mol%) under 596 nm emission (**b**) emission spectra of BMT: xSm^{3+}(x = 0.5,1.0,1.5,2.0 and 2.5 mol%) phosphors under 409 nm excitation wavelength, (**c**) PL excitation and emission spectrum of BMT: yEu^{3+}(y = 1.0 mol%) phosphor under 615 nm emission and 395 nm excitation wavelengths, respectively and (**d**) PL excitation spectrum of BMT: xSm^{3+}/yEu^{3+}(x = 1.5 mol%, y = 1 mol%) phosphor under 615 nm emission wavelength (inset figure shows the overlapped excitation spectrums of singly Sm^{3+} doped and Sm^{3+}/Eu^{3+} co-doped phosphors at emission wavelengths 596 and 615 nm respectively)

Eu^{3+} ion occupies anti-inversion symmetry and gives red emission corresponding to transition $^5D_0 \rightarrow {}^7F_2$ with high colour purity [47].

The PL excitation spectrum of BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1 mol%) co-doped phosphor is shown in Fig. 8(d) in the wavelength range 300–500 nm by keeping λ_{em} = 615 nm. In the spectrum, the highest number of peaks are due to Eu^{3+} ions, and some peaks are due to Sm^{3+} ions. The Sm^{3+} ion peaks at 343, 406, 440, and 480 nm are due to the transitions $^6H_{5/2} \rightarrow {}^4H_{9/2}$, $^6H_{5/2} \rightarrow {}^4F_{7/2}$, $^6H_{5/2} \rightarrow {}^4G_{9/2}$ and $^6H_{5/2} \rightarrow {}^4I_{11/2} + {}^4I_{13/2} + {}^4M_{15/2}$ respectively. Eu^{3+}ion peaks are present at 306, 362, 382, 395, 417, and 465 nm, corresponding to transitions CTB, $^7F_0 \rightarrow {}^5D_4, {}^7F_0 \rightarrow {}^5L_7, {}^7F_0 \rightarrow {}^5L_6, {}^7F_0 \rightarrow {}^5D_3$ and $^7F_0 \rightarrow {}^5D_2$ respectively. The inset of Fig. 8(d) shows the excitation spectrum of BMT: xSm^{3+} (x = 1.5 mol%) and BMT: yEu^{3+} (y = 1 mol%) by keeping the emission wavelengths at 596 and 615 nm, respectively. By taking into account this inset of Fig. 8(d), the emission spectra of BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1, 2, 3, 4, 5, and 6 mol%) co-doped phosphors were recorded at three different excitation wavelengths (381, 409 and 465 nm). Here, 409 nm is the Sm^{3+} ion (sensitizer) excitation wavelength, and 381 and 465 nm are the overlapping wavelengths of Sm^{3+} and Eu^{3+} ions.

Figure 9(a), (b) and (c) show the emission spectra of co-doped BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1, 2, 3, 4, 5 and 6 mol%) phosphors at λ_{ex} = 381, 409, and 465nm respectively. The emission spectra at λ_{ex} = 381 nm show seven peaks. Out of these seven peaks, two peaks at 562 nm ($^4G_{5/2} \rightarrow {}^6H_{5/2}$) and 644 nm ($^4G_{5/2} \rightarrow {}^6H_{7/2}$) correspond to the Sm^{3+} ion transitions, and the other five peaks at 581 ($^5D_0 \rightarrow {}^7F_0$), 594 ($^5D_0 \rightarrow {}^7F_1$), 615 ($^5D_0 \rightarrow {}^7F_2$), 652 ($^5D_0 \rightarrow {}^7F_3$) and 700 nm ($^5D_0 \rightarrow {}^7F_4$) due to the Eu^{3+} ion transitions [48]. The spectra given in Fig. 9(a), (b) and (c) show that with increasing concentration in Eu^{3+} ions, the emission intensity of Eu^{3+}ion peaks increase while Sm^{3+} ion peak intensity decreases. The emission spectra given in Fig. 9(b) were recorded at λ_{ex} = 409 nm. With an increased concentration of Eu^{3+} ions, the emission intensity of Sm^{3+} peaks decreased, and Eu^{3+} ion peaks increased up to 4 mol%. This confirms the energy transfer mechanism of Sm^{3+} ions to Eu^{3+} ions. After a 4 mol% concentration of Eu^{3+} ions, the energy transfer mechanism becomes constant, i.e., it attains saturation. This may be due to the phenomenon of concentration quenching. Figure 9(d) shows the increase in Red to Orange (R/O) emission intensity with different Eu^{3+} ion concentrations at λ_{ex} = 409 nm. The R/O ratio is increasing up to 4 mol% concentration of Eu^{3+} ions. Increasing R/O ratio is a clear indication of energy transfer from Sm^{3+} to Eu^{3+} ions [49]. To explain that energy transfer mechanism via multipolar interaction, Dexter's formula, along with Reisfeld's approximation, was applied as [50].

$$\frac{\eta_0}{\eta} \propto C^{n/3},$$

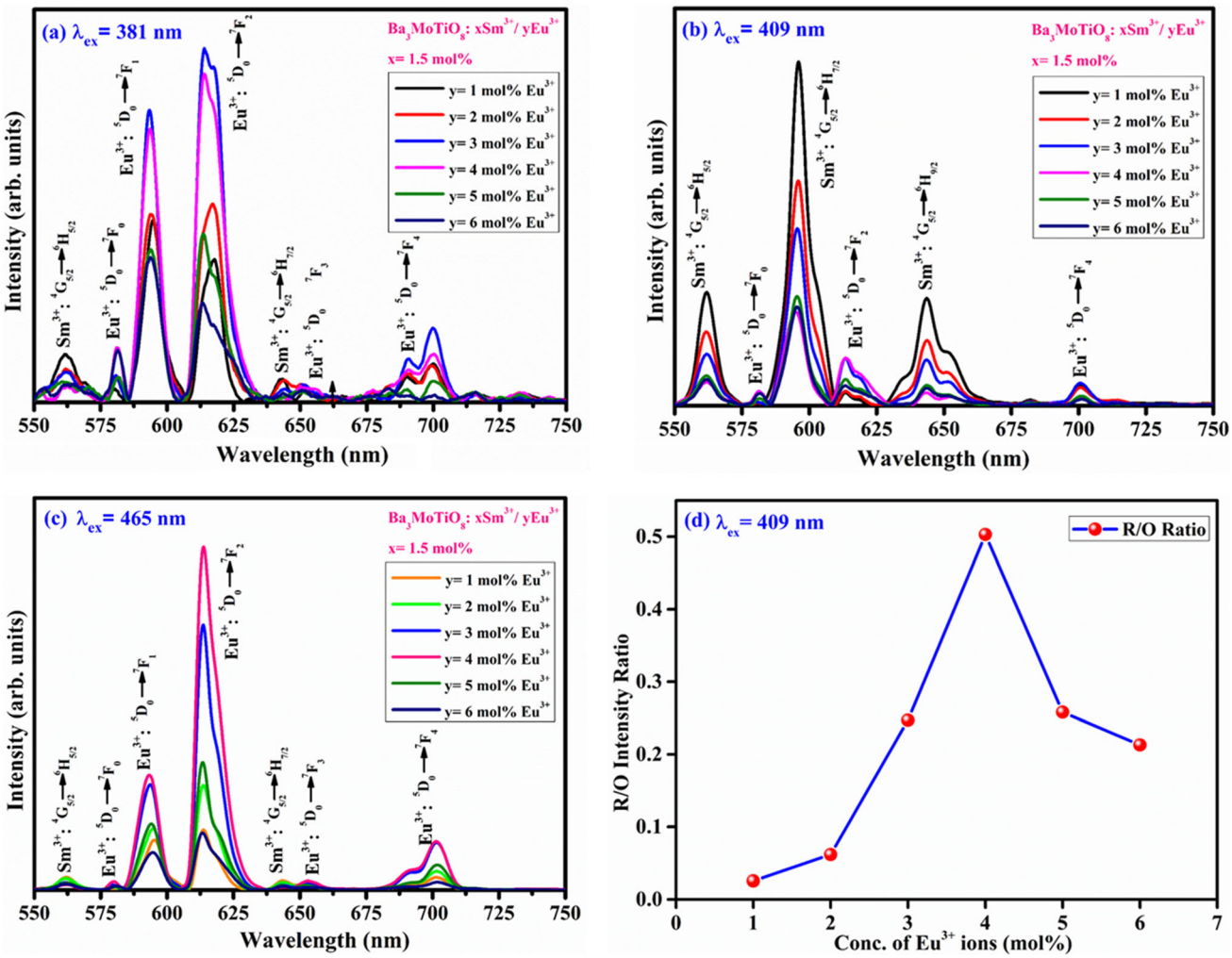


Fig. 9 PL emission spectra of BMT: xSm^{3+}/yEu^{3+} (x= 1.5 mol%, y= 1, 2, 3, 4, 5 and 6 mol%) co-doped phosphors under excitation wavelengths (**a**) 381 nm, (**b**) 409 nm and (**c**) 465 nm respectively and (**d**) plot of R/O ratio vs conc. of Eu^{3+} ions under l_{ex}= 409 nm

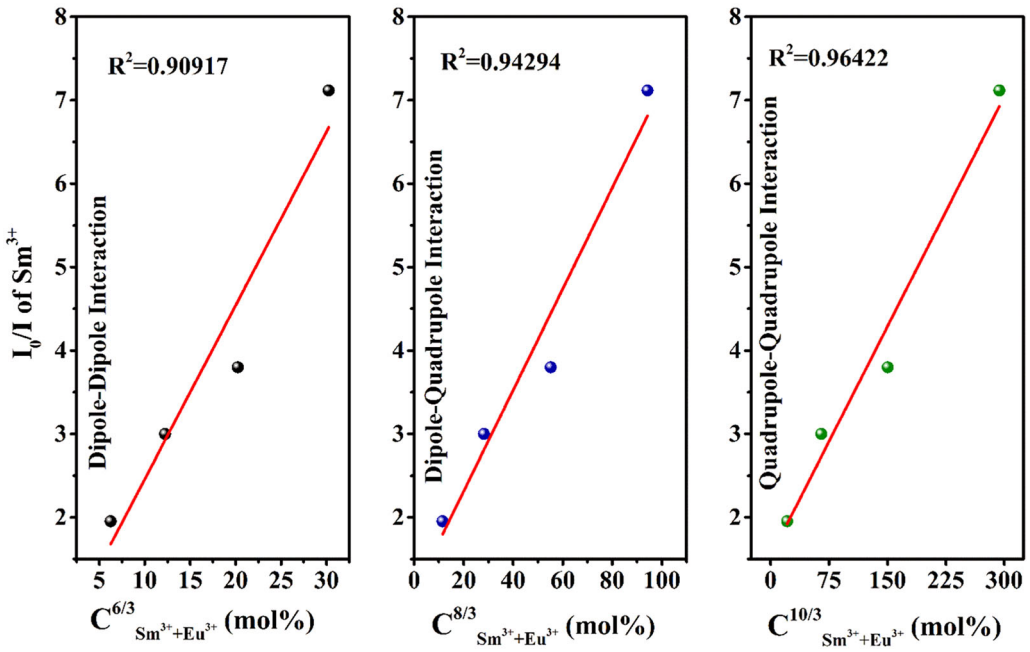


Fig. 10 Variation of I_{SM}/I_{SU} of Sm^{3+} ion w.r.t. $C^{6/3}$, $C^{8/3}$ and $C^{10/3}$ for BMT: xSm^{3+}/yEu^{3+} (x= 1.5mol%, y= 1, 2, 3, 4, 5 and 6 mol%) co-doped phosphors under 409 nm excitation wavelength

where η_0 and η represent the luminescent quantum efficiencies of Sm^{3+} ions without Eu^{3+} ion doping and with Eu^{3+} ion doping, respectively, C is the sum of the concentration of activator (Eu^{3+}) and sensitizer (Sm^{3+}) ions, and n is a number that indicates the type of multipolar interaction (n = 6, 8 and 10 for dipole–dipole, dipole–quadrupole, and quadrupole–quadrupole interaction, respectively). The ratio of luminescence intensities can be used to roughly calculate the ratio of quantum efficiencies [51, 52].

$$\frac{I_{SM}}{I_{SU}} \propto C^{n/3},$$

here I_{SM} is the luminescence emission intensity for the Sm^{3+} ion-doped sample, and I_{SU} is the luminescence emission intensity for the Sm^{3+}/Eu^{3+}co-doped sample. A graph was plotted between I_{SM}/ I_{SU} versus $C^{n/3}$ by taking n = 6, 8, and 10 under λ_{ex} = 409 nm in Fig. 10. The best linear regression was found in the case of n = 10, which shows that quadrupole-quadrupole interaction is responsible for energy transfer mechanism from Sm^{3+} to Eu^{3+} ions [42].

Figure 11 shows the energy level diagram of Sm^{3+} and Eu^{3+} ions to display the energy transfer mechanism between these two ions. There is a very small energy difference between the energy level $^4G_{5/2}$ of the Sm^{3+} ion and the $^5D_0$ level of the Eu^{3+} ion. Due to this, the Sm^{3+} ion can act as an efficient sensitizer for Eu^{3+} ions [53].

3.5 PL decay studies

The PL decay curves of BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1, 2, 3, 4, 5, and 6 mol%) phosphors excited under a 409 nm wavelength have been shown in Fig. 12. The PL decay curves show best fitting for a non-linear bi-exponential function given by [40, 54]-

$$I = I_0 + A_1 exp.\left(\frac{-t}{\tau_1}\right) + A_2 exp.\left(\frac{-t}{\tau_2}\right).$$

where I_0 and I are the luminescence intensities at t = 0 and some moment of time t, respectively. For bi-exponential fitting, A_1 and A_2 are fit constants, while τ_1 and τ_2 are the rapid and slow decay components, respectively. The average lifetime values for bi-exponential curve fitting given by [21, 55].

$$\tau_{avg.} = \frac{A_1\tau_1^2 + A_2\tau_2^2}{A_1\tau_1 + A_2\tau_2}.$$

The average lifetime values of the co-doped BMT phosphors are listed in Table 2 [40, 56, 57]. The lifetime values of the Sm^{3+}/Eu^{3+} co-doped BMT phosphors initially decrease up to 4 mol% of Eu^{3+} ions

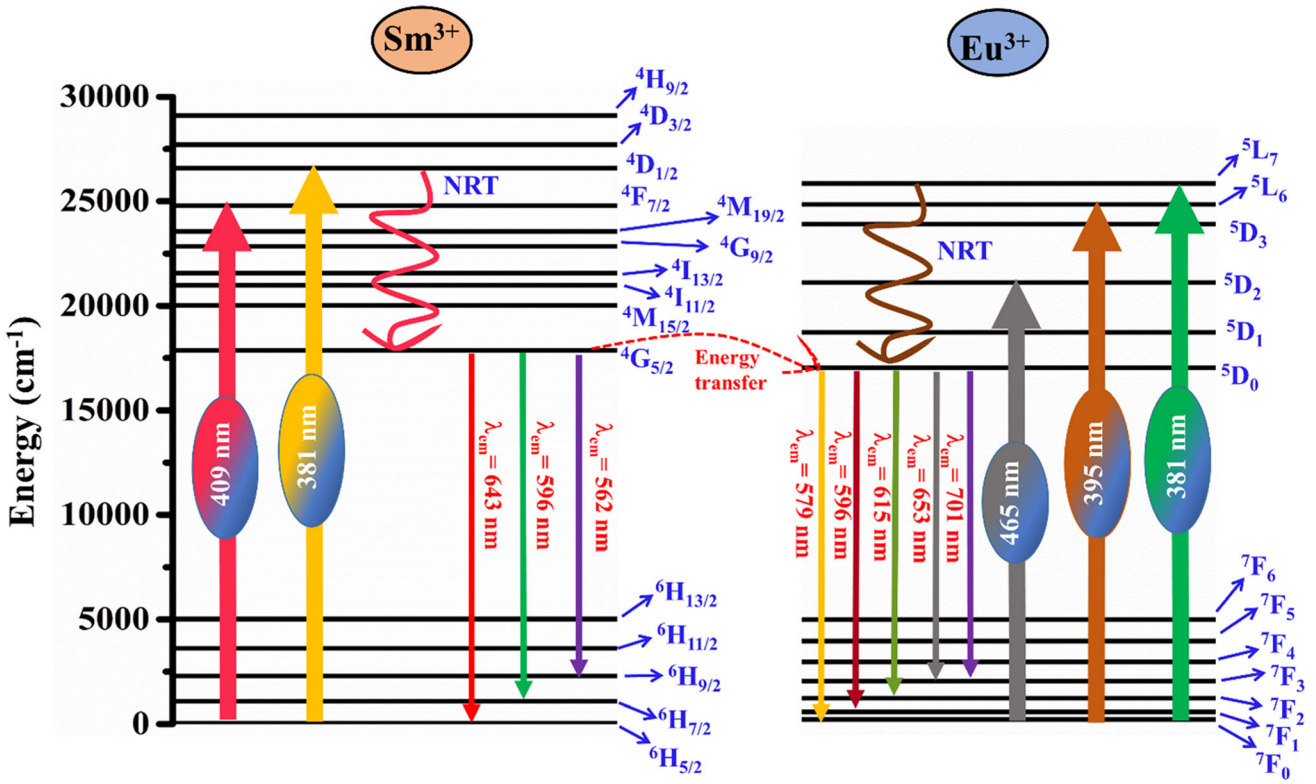


Fig. 11 Schematic energy level diagram of Sm^{3+} and Eu^{3+} ions and their energy transfer mechanism

and reach saturation beyond. This indicates energy transfer from a donor (Sm^{3+} ions) to an acceptor (Eu^{3+}) ion, up to 4 mol% of Eu^{3+} ions in the BMT phosphors. This result is in consonance with the information given by the PL spectral data of Sm^{3+}/Eu^{3+} co-doped BMT phosphors. The PL intensity of the co-doped BMT phosphor increases up to 4 mol% of Eu^{3+} ion concentration and decreases further [47, 56]. After that concentration, energy transfer saturates and results in an intensity decrement in the PL spectra due to the phenomenon of concentration quenching. As a result of it, lifetime decay curve behaviour also changes.

3.6 Temperature-dependent PL studies

Thermal stability is an important parameter for phosphor utilisation in SSL and wLED applications. To check the thermal stability of BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 4 mol%) co-doped phosphor, temperature-dependent PL spectra have been recorded by taking the 409 nm excitation wavelength. Figure 13(a) shows the PL spectral curves by varying the temperature from 303 to 453 K. With variation in temperature, a decrease in the luminescence intensity of the phosphor has been observed without any change in the emission profile. This decrement in intensity may be due to the thermal quenching effect, which increases the non-radiative electronic transitions and results in low luminescence intensity. Up to the 423 K, the phosphor retains approximately 65% of its luminescence intensity, showing the thermal stability of this phosphor. The activation energy of the co-doped phosphor can be calculated using this equation [51].

$$I_T = \frac{I_0}{1 + C\,exp.\left(\frac{-E}{K_B T}\right)}.$$

where I_0 and I_T represent the luminescence intensities at room temperature and at a certain temperature, respectively. Here, C is an equation constant. Figure 13(b) shows a graph plotted between $\ln(I_0/I_T - 1)$ and $1/K_BT$, and linear fitting in this plot gives the value of the slope. The activation energy (ΔE = 0.306) calculated from the slope was compared with the

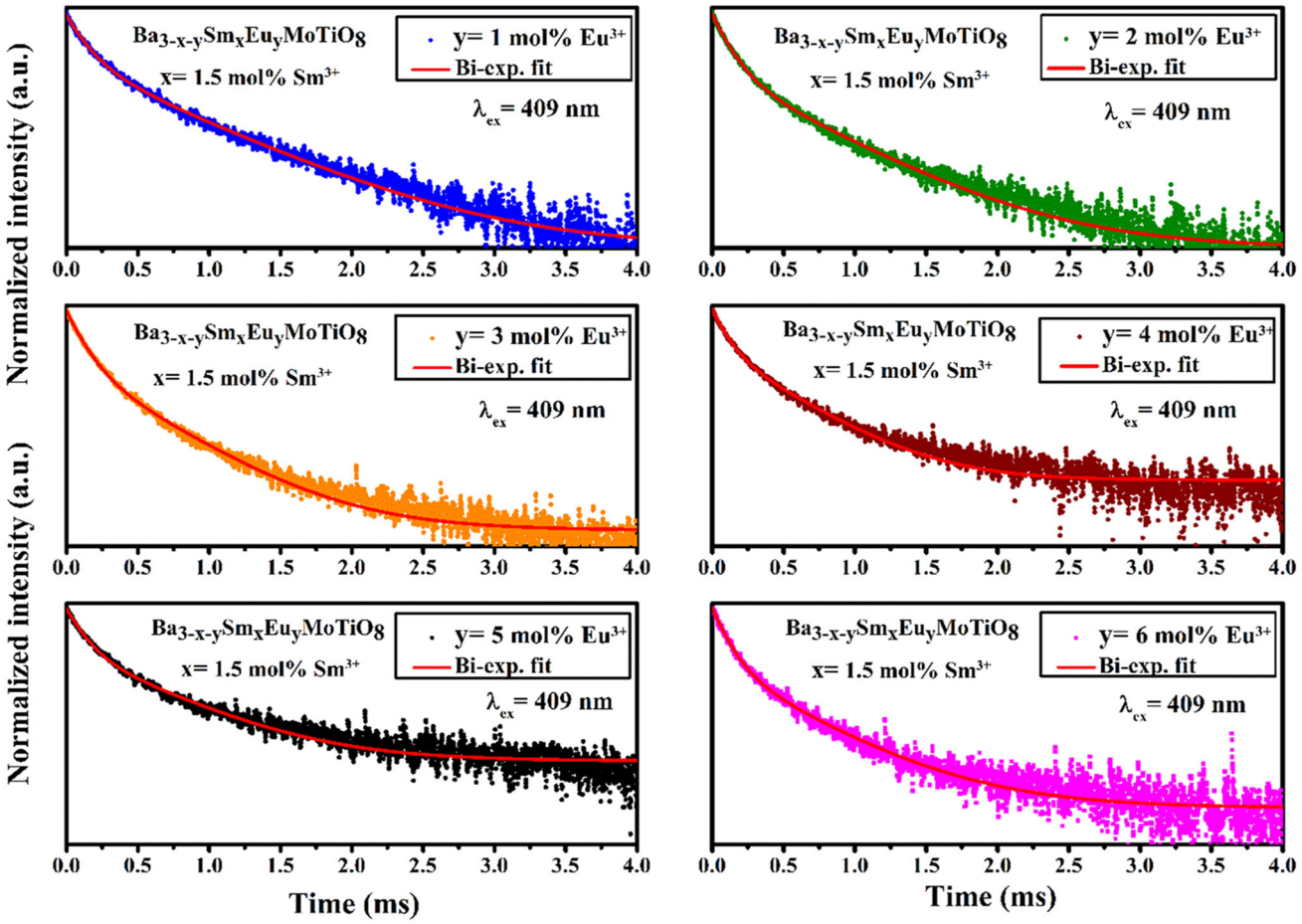


Fig. 12 PL lifetime curves of BMT: xSm^{3+}/yEu^{3+} (x=1.5 mol% and y=1, 2, 3, 4, 5 and 6 mol%) phosphors under 409 nm excitation

Table 2 Lifetime decay values of BMT: xSm^{3+}/yEu^{3+}(x = 1.5 mol%, y = 1,2,3,4,5 and 6 mol%) co-doped phosphors for λ_{ex} = 409 nm

Sample ID (Sm^{3+}/Eu^{3+}series) (mol%)	Lifetime values (ms)
y = 1	0.60 References [40, 56, 57].
y = 2	0.448
y = 3	0.257
y = 4	0.205
y = 5	0.273
y = 6	0.311

previously reported work and showed the good thermal stability of the co-doped phosphor [58].

3.7 CIE performance

Using PL emission spectral data, CIE coordinates were determined for the optimised single- and co-doped BMT phosphors, as shown in Fig. 14. The CIE coordinates for the optimised single Sm^{3+} ion (1.5 mol% concentration) doped BMT phosphor were (0.582, 0.416) under the 409 nm excitation wavelength, which lie in the orange- reddish region. CIE co-ordinates for the BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 4 mol%) phosphor were calculated to be (0.632, 0.36), (0.598, 0.40) and (0.650, 0.349) under 381 nm, 409 nm, and 465 nm excitation wavelengths,

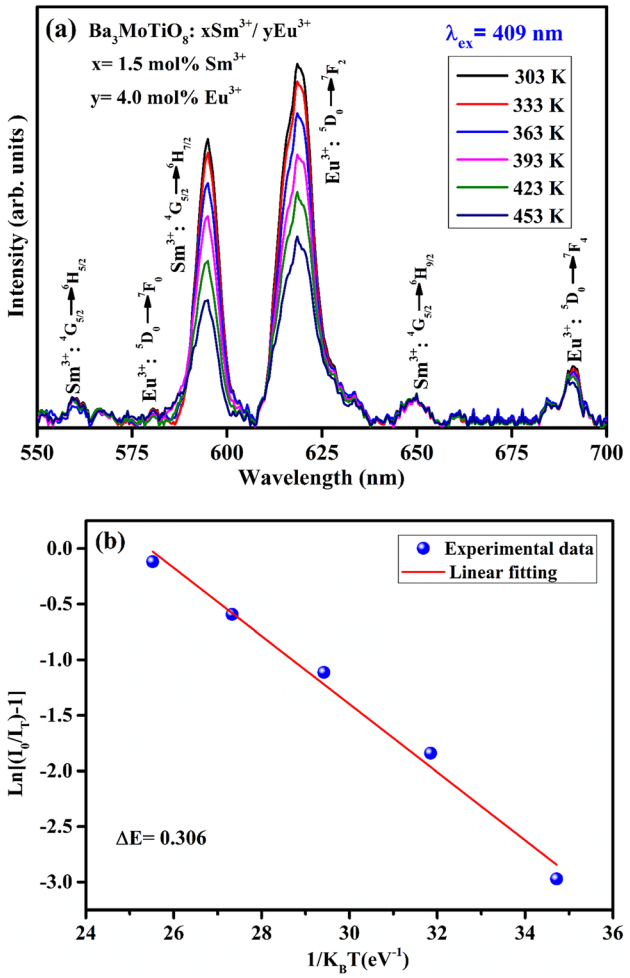


Fig. 13 Temperature-dependent PL emission spectra (**a**) of BMT: xSm^{3+}/yEu^{3+} (x= 1.5 mol%, y= 1, 2, 3, 4, 5 and 6 mol%) co-doped phosphors under excitation wavelengths 409 nm and (**b**) graph of $Ln(I_0/I_T-1)$ vs. $1/K_BT$ for activation energy

respectively. All the CIE coordinates for co-doped phosphors lie in the pure red region.

4 Conclusion

A series of singly Sm^{3+} doped BMT: xSm^{3+} (x = 0.5, 1, 1.5, 2 and 2.5 mol%) and co-doped BMT: xSm^{3+}/yEu^{3+} (x = 1.5 mol%, y = 1, 2, 3, 4, 5 and 6 mol%) phosphors were successfully synthesized by combustion method. The single-phase confirmation of these singly doped and co-doped phosphors was done by XRD patterns. All the phosphors show sharp XRD peaks without any impurities. Through SEM, morphological analysis has been performed, and particle size was calculated to be in the micrometer range. All the elements were uniformly distributed over the surface of the crystal, and that has been confirmed by EDX. FT-IR spectra recorded for an undoped, single Sm^{3+} doped, and Sm^{3+}/Eu^{3+} co-doped phosphors show similar vibrational and absorption bands. The optical band gap obtained from the DRS study lies in the range of 3.0–3.5 eV for all the concentrations of Sm^{3+} and Sm^{3+}/Eu^{3+} co-doped phosphors. PL spectra recorded under the 409 nm excitation wavelength for the singly Sm^{3+} ion doped phosphors show reddish-orange emission at the 596 nm wavelength, corresponding to transition $^4G_{5/2} \rightarrow {}^6H_{7/2}$. Sharp red emission at 615 nm corresponding to transition $^5D_0 \rightarrow {}^7F_2$ is visible in PL spectra for co-doped BMT phosphors that were taken at three distinct excitation wavelengths (409, 381, and 465 nm). The intensity of the Eu^{3+} peaks increases with an increase in Eu^{3+} ion concentration, while the Sm^{3+} peaks decrease under the 409 nm excitation wavelength, according to PL spectral analysis of co-doped phosphors. This shows the energy transfer mechanism from Sm^{3+} ions to Eu^{3+} ions in the co-doped phosphors. Quadrupole–quadrupole interaction is the basis for this energy transfer process, according to Dexter's theory applied to it. CIE coordinates calculated for optimized single Sm^{3+} ion-doped phosphors lie in the reddish–orange region, and for co-doped phosphors lie in the red region, demonstrating the energy transfer induced colour tunable behaviour from the reddish-orange to red region, respectively. These studies show the application of prepared phosphors in orange and red color-emitting LED devices.

Acknowledgements

The Delhi Technological University provided the facilities required to enable the current research endeavour, for which the authors are quite grateful. Additionally, we would like to express our gratitude to Dr. Manju Srivastava and Dr. Shyama Prasad of the Chemistry Department at the Dayalbagh Educational Institute of the Deemed University of Agra for providing all the assistance necessary to take FESEM images of the samples and record the EDAX data.

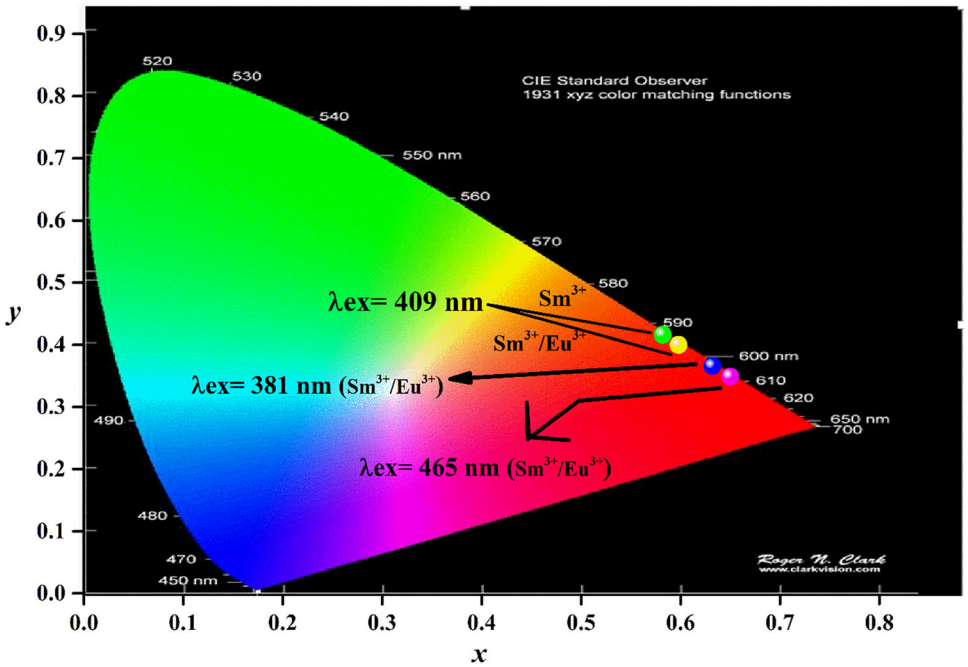


Fig. 14 CIE plot of optimised single Sm^{3+} doped (under 409 nm excitation) and Sm^{3+}/Eu^{3+} co-doped phosphors (under 381, 409, and 465 nm excitation), respectively

Author contributions

PR: experimental, calculations, analysis, writing. ASR: work plan, validation of results, manuscript corrections.

Funding

One of the author Pooja Rohilla is thankful to the Delhi Technological University for providing Senior Research Fellowship (SRF) to support this research work.

Data availability

The data that support the findings of this study are available on request from the corresponding author Prof. A.S. Rao. The data are not publicly available due to restrictions, e.g., their containing information that could compromise the privacy of research participants.

Declarations

Competing interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. X. Zhou, Z. Zhang, Y. Wang, Ce^{3+} and Tb^{3+} singly- and co-doped $MgGd_4Si_3O_{13}$ for ultraviolet light emitting diodes and field emission displays. J. Mater. Chem. C **3**, 3676–3683 (2015). https://doi.org/10.1039/c4tc02807d
2. Z. Zhu, Y. Zhang, Y. Qiao, H. Liu, D. Liu, Luminescence properties of $Ce^{3+}/Tb^{3+}/Sm^{3+}$ co-doped $CaO–SiO_2–B_2O_3$ glasses for white light emitting diodes. J. Lumin. **134**, 724–728 (2013). https://doi.org/10.1016/j.jlumin.2012.07.003
3. C.H. Chiang, H.H. Su, Y.C. Fang, S.Y. Chu, Effects of argon sintering atmosphere on luminescence characteristics of $Ca_6BaP_4O_{17}:Sm^{3+}$ phosphors. Ceram. Int. **44**, 6278–6284 (2018). https://doi.org/10.1016/j.ceramint.2018.01.016
4. Z. Jia, C. Yuan, Y. Liu, Strategies to approach high performance in Cr^{3+} -doped phosphors for high-power NIR-LED light sources. Light Sci. Appl. **9**, 86 (2020). https://doi.org/10.1038/s41377-020-0326-8
5. W.U. Khan, L. Zhou, Q. Liang, X. Li, J. Yan, N.U. Rahman, L. Dolgov, S.U. Khan, J. Shi, M. Wu, Luminescence enhancement and energy transfers of Ce^{3+} and Sm^{3+} in $CaSrSiO_4$ phosphor. J. Mater. Chem. C **6**, 7612–7618 (2018). https://doi.org/10.1039/c8tc02143k
6. M. Janulevicius, P. Marmokas, M. Misevicius, J. Grigorjevaite, L. Mikoliunaite, S. Sakirzanovas, A. Katelnikovas,

Luminescence and luminescence quenching of highly efficient $Y_2Mo_4O_{15}:Eu^{3+}$ phosphors and ceramics. Sci. Rep. **6**, 1–12 (2016). https://doi.org/10.1038/srep26098

7. S.E. Brinkley, N. Pfaff, K.A. Denault, Z. Zhang, H.T. Hintzen, R. Seshadri, S. Nakamura, S.P. Denbaars, Robust thermal performance of $Sr_2Si_5N_8:Eu^{2+}$: an efficient red emitting phosphor for light emitting diode based white lighting. Appl. Phys. Lett. **99**, 24116 (2011). https://doi.org/10.1063/1.3666785
8. G. Annadurai, S.M.M. Kennedy, Synthesis and photoluminescence properties of $Ba_2CaZn_2Si_6O_{17}:Eu^{3+}$ red phosphors for white LED applications. J. Lumin. **169**, 690–694 (2016). https://doi.org/10.1016/j.jlumin.2015.03.007
9. A.K. Vishwakarma, M. Jayasimhadri, Pure orange color emitting Sm^{3+} doped $BaNb_2O_6$ phosphor for solid-state lighting applications. J. Lumin. **176**, 112–117 (2016). https://doi.org/10.1016/j.jlumin.2016.03.025
10. T. Srikanth, D.V.K. Reddy, K.S. Rudramamba, S. Taherunnisa, N.M. Reddy, T.L. Viveka, M.R. Reddy, Red light component tuning by n-UV/blue light excitations in Sm^{3+}/Eu^{3+} co-doped Y_2O_3–Al_2O_3–Bi_2O_3–B_2O_3–SiO_2 glasses for W-LED applications. Opt. Mater. (Amst.). 134, 13148 (2022). https://doi.org/10.1016/j.optmat.2022.113148
11. S. Kaur, N. Deopa, A. Prasad, R. Bajaj, A.S. Rao, Intense green emission from Tb^{3+} ions doped zinc lead alumino borate glasses for laser and w-LEDs applications. Opt. Mater. (Amst). **84**, 318–323 (2018). https://doi.org/10.1016/j.optmat.2018.07.020
12. N. Deopa, M.K. Sahu, P.R. Rani, R. Punia, A.S. Rao, Realization of warm white light and energy transfer studies of Dy^{3+}/Eu^{3+} co-doped Li_2O–PbO–Al_2O_3–B_2O_3 glasses for lighting applications. J. Lumin. **222**, 117166 (2020). https://doi.org/10.1016/j.jlumin.2020.117166
13. N. Deopa, B. Kumar, M.K. Sahu, P.R. Rani, A.S. Rao, Effect of Sm^{3+} ions concentration on borosilicate glasses for reddish orange luminescent device applications. J. Non Cryst. Solids. **513**, 152–158 (2019). https://doi.org/10.1016/j.jnoncrysol.2019.03.025
14. K.N. Kumar, L. Vijayalakshmi, J.S. Kim, Enhanced red luminescence quantum yield from $Gd^{3+}/Eu^{3+}:CaLa_2ZnO_5$ phosphor spheres for photonic applications. Mater. Res. Bull. **103**, 234–241 (2018). https://doi.org/10.1016/j.materresbull.2018.03.045
15. V. Balaram, Rare earth elements: a review of applications, occurrence, exploration, analysis, recycling, and environmental impact. Geosci. Front. **10**, 1285–1303 (2019). https://doi.org/10.1016/j.gsf.2018.12.005
16. Y. Tian, Development of phosphors with high thermal stability and efficiency for phosphor-converted LEDs. J. Solid State Lighting. **1**, 11 (2014). https://doi.org/10.1186/s40539-014-0011-8
17. D. Chiriu, L. Stagi, C.M. Carbonaro, P.C. Ricci, Strength and weakness of rare earths based phosphors: strategies to replace critical raw materials. Phys. Status Solidi Curr. Top. Solid State Phys. **13**, 989–997 (2016). https://doi.org/10.1002/pssc.201600116
18. L. Su, Y. Liu, G. Li, Z. Wu, G. Cai, L. Ma, X. Wang, Multicolor emission leading by energy transfer between Dy^{3+} and Eu^{3+} in wolframite $InNbTiO_6$. J. Lumin. **227**, 17578 (2020). https://doi.org/10.1016/j.jlumin.2020.117578
19. Y. Tang, S. Zhou, X. Yi, D. Hao, X. Shao, J. Chen, The characterization of Ce/Pr-doped YAG phosphor ceramic for the white LEDs. J. Alloys Compd. **745**, 84–89 (2018). https://doi.org/10.1016/j.jallcom.2018.02.129
20. S. Kaur, M. Jayasimhadri, A.S. Rao, A novel red emitting Eu^{3+}doped calcium aluminozincate phosphor for applications in w-LEDs. J. Alloys Compd. **697**, 367–373 (2017). https://doi.org/10.1016/j.jallcom.2016.12.150
21. S. Lou, P. Zhang, Y. Chen, Q. Xiong, K. Qiu, Synthesis and luminescence enhancement of $CaY_{0.6}(MoO4)_{1.9}:Eu^{3+}$ red phosphors by Sm^{3+} co-doping. Ceram. Int. **47**, 10174–10184 (2021). https://doi.org/10.1016/j.ceramint.2020.12.166
22. P.K. Vishwakarma, A. Bahadur, A. Maurya, S.B. Rai, Large enhancement in upconverted green emission intensity from Ho^{3+} / Yb^{3+} co-doped $Y_2Ti_2O_7$ phosphor in the presence of zn 2 +. Mater. Res. Bull. **115**, 219–226 (2019). https://doi.org/10.1016/j.materresbull.2019.03.029
23. P. Rohilla, A.S. Rao, Linear and non-linear photoluminescence studies of Ho^{3+}/Yb^{3+} co-doped titanate phosphors for photonic applications. J. Alloys Compd. **928**, 167156 (2022). https://doi.org/10.1016/j.jallcom.2022.167156
24. P. Du, J.S. Yu, Eu^{3+}-activated La_2MoO_6–La_2WO_6 red-emitting phosphors with ultrabroad excitation band for white light-emitting diodes. Sci. Rep. **7**, 1–10 (2017). https://doi.org/10.1038/s41598-017-12161-5
25. P. Rohilla, A.S. Rao, Synthesis optimisation and efficiency enhancement in Eu^{3+} doped barium molybdenum titanate phosphors for w-LED applications. Mater. Res. Bull. **150**, 111753 (2022). https://doi.org/10.1016/j.materresbull.2022.111753
26. S. Kaur, A.S. Rao, M. Jayasimhadri, Spectroscopic and photoluminescence characteristics of Sm^{3+} doped calcium aluminozincate phosphor for applications in w-LED. Ceram. Int. **43**, 7401–7407 (2017). https://doi.org/10.1016/j.ceramint.2017.02.129
27. M. Fu, Q. Wu, C. Gu, S. Hu, S. Lu, B. Wang, Y. Hong, Y. Wang, Energy transfer in Sm^{3+} /Eu^{3+} doped Li Gd(BO) orange-red phosphor. Vacuum. **214**, 11247 (2023). https://doi.org/10.1016/j.vacuum.2023.112147

28. P.K. Vishwakarma, S.B. Rai, A. Bahadur, Intense red and green emissions from Ho^{3+}/Yb^{3+} co-doped sodium gadolinium molybdate nano-phosphor: effect of calcination temperature and intrinsic optical bistability. Mater. Res. Bull. **133**, 111041 (2021). https://doi.org/10.1016/j.materresbull.2020.111041
29. X. Yuhang, Z. Xiangyu, Z. Hongbo, Z. Mengjie, M. Shuo, S. Chunhui, J. Shao, Luminescence properties of Eu^{3+} doped $BaMoO_4$ transparent glass ceramics. J. Non Cryst. Solids. **500**, 243–248 (2018). https://doi.org/10.1016/j.jnoncrysol.2018.08.007
30. F. Xu, L. Zhong, Y. Xu, S. Feng, C. Zhang, F. Zhang, G. Zhang, Highly efficient flame-retardant kraft paper. J. Mater. Sci. **54**, 1884–1897 (2019). https://doi.org/10.1007/s10853-018-2911-2
31. X. Zhao, J. Zhang, L. Shi, M. Xian, C. Dong, S. Shuang, Folic acid-conjugated carbon dots as green fluorescent probes based on cellular targeting imaging for recognizing cancer cells. RSC Adv. **7**, 42159–42167 (2017). https://doi.org/10.1039/c7ra07002k
32. C.P. Huang, M.A. Rather, C.T. Wu, R. Loganathan, Y.H. Ju, K.L. Lin, J.I. Chyi, K.Y. Lai, Crystal transformation of cubic BN nanoislands to rhombohedral BN sheets on AlN for deep-UV light-emitting diodes, ACS Appl. Nano Mater. **3**, 5285–5290 (2020). https://doi.org/10.1021/acsanm.0c00681
33. B. Viswanathan, S. Murugesan, A. Ariharan, K.S. Lakhi, Hetero atom substituted carbon—potential hydrogen storage materials. Adv. Porous Mater. **1**, 122–128 (2013). https://doi.org/10.1166/apm.2013.1008
34. L. Zhang, B. Sun, C. Shao, F. Zhen, S. Wei, W. Bu, Q. Yao, Z. Jiang, H. Chen, Preparation, band-structure and luminescence of double perovskite Ba_2MgMoO_6:Eu^{3+} orange-red phosphor for white LEDs. Ceram. Int. **44**, 17305–17312 (2018). https://doi.org/10.1016/j.ceramint.2018.06.192
35. W. Yuan, R. Pang, S. Wang, T. Tan, C. Li, C. Wang, H. Zhang, Enhanced blue-light excited cyan-emitting Bi^{3+} phosphors for AC-LEDs via defect modulation. Light: Sci App. **11**, 184 (2022). https://doi.org/10.1038/s41377-022-00868-8
36. D. Jiang, L. Geng, S. Zhou, Y. Wang, Photoluminescence properties and energy transfer in the Sm^{3+} and Eu^{3+} co-doped $Ca_3Bi(PO_4)_3$ red phosphor. Inorg. Chem. Commun. **142**, 109668 (2022). https://doi.org/10.1016/j.inoche.2022.109668
37. A.S. Asyikin, M.K. Halimah, A.A. Latif, M.F. Faznny, S.N. Nazrin, Physical, structural and optical properties of bio-silica borotellurite glass system doped with samarium oxide nanoparticles. J. Non Cryst. Solids. **529**, 119777 (2020). https://doi.org/10.1016/j.jnoncrysol.2019.119777
38. S.P. Khambule, S.V. Motloung, T.E. Motaung, L.F. Koao, R.E. Kroon, M.A. Malimabe, Results in optics tuneable blue to orange phosphor from Sm^{3+} doped $ZnAl_2O_4$ nanomaterials. Results Opt. **9**, 100280 (2022). https://doi.org/10.1016/j.rio.2022.100280
39. C. Ta, H. Wen, Z. Mo, Y. Huo, Z. Mu, E.A. Dawi, L. Luo, D. Abdul, Preparation and luminescence investigations of garnet-based $Y_2Ca_2Ga_3VO_{12}$: RE (RE = Eu^{3+}, Sm^{3+}, and Dy^{3+}) phosphors. Mater. Res. Bull. **164**, 112254 (2023). https://doi.org/10.1016/j.materresbull.2023.112254
40. S. Dai, D. Zhao, R. Zhang, L. Jia, Q. Yao, Enhancing luminescence intensity and improving thermostability of red phosphors $Li_3Ba_2La_3(WO_4)_8$_Eu^{3+} by co-doping with Sm^{3+} ions. J. Alloys Compd. **891**, 161973 (2021). https://doi.org/10.1016/j.jallcom.2021.161973
41. J. Yang, H. Xiong, J. Dong, C. Yang, S. Gan, L. Zou, Facile hydrothermal synthesis and luminescent properties of Sm^{3+}/Eu^{3+} codoped $GdPO_4$ phosphors. J. Phys. Chem. Solids **111**, 355–363 (2017). https://doi.org/10.1016/j.jpcs.2017.08.031
42. M.K. Sahu, M. Jayasimhadri, K. Jha, B. Sivaiah, A.S. Rao, D. Haranath, Synthesis and enhancement of photoluminescent properties in spherical shaped Sm^{3+}/Eu^{3+} co-doped $NaCaPO_4$ phosphor particles for w-LEDs. J. Lumin. **202**, 475–483 (2018). https://doi.org/10.1016/j.jlumin.2018.06.002
43. J. Gou, D. Zhang, B. Yu, J. Wang, S. Liu, The photoluminescence behaviors of a novel reddish orange emitting phosphor $CaIn_2O_4$: Sm^{3+} Codoped with Zn^{2+} or Al^{3+} Ions. J. Nanomater. **2**, 1-5 (2015). https://doi.org/10.1155/2015/969724
44. H. Tang, X. Zhang, L. Cheng, L. Jiang, X. Mi, Q. Liu, Photoluminescence properties and energy transfer mechanisms of. J. Lumin. **214**, 116532 (2019). https://doi.org/10.1016/j.jlumin.2019.116532
45. X. Gao, W. Zhang, X. Wang, X. Huang, H. Du, J. Lin, Enhancing the photoluminescence performance of $Ca_5(PO_4)_2SiO_4$: Re^{3+} (re = Eu, Sm) phosphors with A^{3+} (A = La, Bi) codoping and white light-emitting diode application. Ceram. Int. **48**, 13080–13089 (2022). https://doi.org/10.1016/j.ceramint.2022.01.184
46. M.N. Huang, Y.Y. Ma, X.Y. Huang, S. Ye, Q.Y. Zhang, S.A., The luminescence properties of Bi^{3+} sensitized Gd_2MoO_6: RE^{3+} (RE = eu or sm) phosphors for solar spectral conversion. Mol. Biomol. Spectrosc. **115**, 767–771 (2013). https://doi.org/10.1016/j.saa.2013.06.111
47. M. Yan, G. Liu, J. Wen, Y. Wang, Blue-light-excited Eu^{3+}/Sm^{3+} co-doped $NaLa(MoO4)_2$ phosphors: synthesis, characterizations and red emission enhancement for WLEDs. Materials. **11**, 1090 (2018). https://doi.org/10.3390/ma11071090

48. V.A. Online, L. Wu, Y. Bai, L. Wu, H. Yi, Y. Kong, Y. Zhang, J. Xu, Sm^{3+} and Eu^{3+} codoped SrBi B O : a red-emitting phosphor with improved thermal stabilityR.S.C. Advances, **7**, 1146–1153 (2017). https://doi.org/10.1039/C6RA26752A
49. A.S. Ravita, Rao, Effective sensitization of Eu^{3+} visible red emission by Sm^{3+} in thermally stable potassium zinc alumino borosilicate glasses for photonic device applications. J. Lumin. **244**, 118689 (2022). https://doi.org/10.1016/j.jlumin.2021.118689
50. N. Liu, L. Mei, L. Liao, J. Fu, D. Yang, High Thermal stability apatite Sm^{3+} for white light emission: synthesis, structure, luminescence properties and energy transfer. Sci. Rep. **8**, 1–9 (2019). https://doi.org/10.1038/s41598-019-51915-1
51. Y. Ding, N. Guo, M. Zhu, W. Lv, R. Ouyang, Y. Miao, Luminescence and temperature sensing abilities of zincate phosphors co-. Mater. Res. Bull. **129**, 110869 (2020). https://doi.org/10.1016/j.materresbull.2020.110869
52. S. Ling, F.B. Xiong, W.B. Yang, H.F. Lin, W.Z. Zhu, Novel Sm^{3+}/Eu^{3+} co-doped Sr_7Sb_2 O_{12} red-emitting phosphor for white LED. Inorg. Chem. Commun **150**, 110365 (2023). https://doi.org/10.1016/j.inoche.2022.110365
53. P.K. Pandey, P. Dixit, V. Chauhan, P.C. Pandey, Luminescence properties and energy transfer studies in thermally stable Bi_2O_3_ Sm^{3+}, Eu^{3+} phosphor. J. Alloys Compd. **952**, 169911 (2023). https://doi.org/10.1016/j.jallcom.2023.169911
54. D. Liu, G. Li, P. Dang, Q. Zhang, Y. Wei, L. Qiu, M.S. Molokeev, H. Lian, M. Shang, J. Lin, Highly efficient Fe^{3+}-doped $A_2BB'O_6$ (A = Sr^{2+}, infrared-emitting phosphors for spectroscopic analysis. Light Sci. Appl. **11**, 112 (2022). https://doi.org/10.1038/s41377-022-00803-x
55. G. Yuan, R. Cui, J. Zhang, X. Zhang, X. Qi, C. Deng, Photoluminescence evolution and high thermal stability of orange red-emitting $Ba_{3-x}Sr_xZnNb_2O_9$: Eu^{3+} phosphors. J. Solid State Chem. **303**, 122447 (2021). https://doi.org/10.1016/j.jssc.2021.122447
56. R. Cui, X. Wang, Q. Chen, C. Deng, Photoluminescence properties and energy transfer of $ZnNb_2O_6$: Eu_{3+}, Sm_{3+} orange–red phosphors for WLEDs. Optik (Stuttg) **261**, 169116 (2022)
57. X. Zhang, R. Cui, K. Guo, M. Zhang, J. Zhang, C. Deng, Luminescence properties of Ca_2 $GdNbO_6$: Sm^{3+}, Eu^{3+} red phosphors with high quantum yield and excellent thermal stability for WLEDs. Ceram. Int. **49**, 15402 (2023). https://doi.org/10.1016/j.ceramint.2023.01.125
58. M. Kumar, A.S. Rao, S. Kaur, Downshifting analysis of Sm^{3+}/Eu^{3+} co-doped LiBiAlBSi glasses for red emission element of white LEDs. Chem. Phys. Lett. **788**, 139303 (2022). https://doi.org/10.1016/j.cplett.2021.139303

Publisher's Note Springer nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/373296972

Enhanced solar cell efficiency: copper zinc tin sulfide absorber thickness and defect density analysis

Article *in* Journal of Materials Science: Materials in Electronics · August 2023

DOI: 10.1007/s10854-023-11125-y

CITATIONS
0

READS
107

12 authors, including:

Devendra Kc
UiT - The Arctic University of Norway, campus Narvik
28 PUBLICATIONS **277** CITATIONS
SEE PROFILE

Deb Kumar Shah
Chonbuk National University
37 PUBLICATIONS **348** CITATIONS
SEE PROFILE

Khim Bahadur Khattri
Kathmandu University
23 PUBLICATIONS **212** CITATIONS
SEE PROFILE

M. Shaheer Akhtar
Chonbuk National University
331 PUBLICATIONS **7,785** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Organic Solar Cell View project

Fabrication of sensors based on Functional nanomaterials View project

All content following this page was uploaded by Deb Kumar Shah on 23 August 2023.

The user has requested enhancement of the downloaded file.

J Mater Sci: Mater Electron (2023) 34:1699

Enhanced solar cell efficiency: copper zinc tin sulfide absorber thickness and defect density analysis

K. C. Devendra[1], Deb Kumar Shah[2], Subhash Kumar[3], Nawraj Bhattarai[4], Dipak Raj Adhikari[5], Khim B. Khattri[6], M. Shaheer Akhtar[2,7,8,*], Ahmad Umar[9,10], Ahmed A. Ibrahim[9], Mohsen A. M. Alhamami[9], Sotirios Baskoutas[11], and O.-Bong Yang[2,7,8,*]

[1] *Lebesby Kommune, Lebesby, Norway*
[2] *School of Semiconductor and Chemical Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea*
[3] *Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India*
[4] *Department of Mechanical and Aerospace Engineering, InstituteofEngineering, Tribhuvan University, Kathmandu 44600, Nepal*
[5] *Graduate School of Science and Technology, Mid-West University, Birendranagar, Nepal*
[6] *School of Science, Kathmandu University, Dhulikhel, Kavre 44600, Nepal*
[7] *Graduate School of Integrated Energy-AI, Jeonbuk National University, Jeonju 54896, Korea*
[8] *New and Renewable Energy Materials Development Center (NewREC), Jeonbuk National University, Jeonbuk 56332, Korea*
[9] *Department of Chemistry, College of Science and Arts and Promising Centre for Sensors and Electronic Devices (PCSED), Najran University, Najran 11001, Kingdom of Saudi Arabia*
[10] *Department of Materials Science and Engineering, The Ohio State University, Columbus, OH 43210, USA*
[11] *Department of Materials Science, University of Patras, 26500 Patras, Greece*

Received: 11 April 2023
Accepted: 7 August 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature, 2023

ABSTRACT

Copper zinc tin sulfide solar cell (CZTS), Cu_2ZnSnS_4-based solar cells have shown promising conversion efficiency because of their ease of variation in configurations. In this work, the architecture of a ZnO–Al/i–ZnO/n–CdS/CZTS/Mo solar cell was optimized by using Silvaco Atlas simulation software. In this simulation study, the thickness and defect density of the CZTS layer has been varied to achieve the highest efficiency of 26.58%, with Isc = 36.64 A and Voc = 0.909 V at a defect density of 1.8×10^{12} cm^{-3}. Increase in the layer thickness of CZTS improves the photon absorption and cell efficiency. This study has evidenced the impact of defect density on the absorber layer, including photo-generation rate, recombination rate, and solar cell efficiency. By optimizing the device parameters, it has achieved a fill factor of 79.74% under AM 1.5 illumination, demonstrating the potential for low-cost, highly efficient CZTS solar cells.

K. C. Devendra and Deb Kumar Shah have contributed equally to this work.

Address correspondence to E-mail: shaheerakhtar@jbnu.ac.kr; obyang@jbnu.ac.kr

https://doi.org/10.1007/s10854-023-11125-y
Published online: 22 August 2023

1 Introduction

In recent decades, human energy consumption has dramatically increased due to rapid scientific and technological advancements. This increase is primarily fueled by our reliance on finite fossil fuels, such as coal, oil, and natural gas, to power our homes, businesses, and transportation infrastructure [1, 2]. However, burning these fuels produces large volumes of carbon dioxide and other greenhouse gases that contribute to global warming [3–7]. Consequently, there has been a global push to transition to cleaner, energy sources such as solar, wind, hydropower, capacitors, fuel cell, batteries, hydrogen, etc., which are abundant, widely accessible, and have a significantly smaller carbon footprint [8–20]. To this end, researchers have focused on developing new solar energy systems that are highly efficient, require less investment, and emit less pollution [21–23]. Among these, thin-film solar cells have generated immense interest among researchers due to their promising potential as a renewable energy technology. In particular, the development of highly efficient absorber materials for solar energy has been a key area of innovation [24–26].

Currently, chalcopyrite Cu(In,Ga) $(S,Se)_2$(CIGS) and CdTe solar cells have high cell efficiencies of 22.6 [27] and 22.1% [28], respectively. However, the toxicity of Cadmium (Cd) and the constrained supply of Indium (In) have raised concerns. Therefore, the use of abundant raw materials is critical to ensure the economic fabrication of solar cells [29]. In recent years, the absorber layer of solar cells constructed from materials such as copper–zinc–tin–sulfide (Cu_2ZnSnS_4 or CZTS), copper–zinc–tin–selenide ($Cu_2ZnSnSe_4$ or CZTSe), and copper–zinc–tin–(sulfur/selenium) ($Cu_2ZnSn(S,Se)_4$ or CZTSSe) have gained interest as potential replacements for conventional solar cells [30]. Compared to other materials such as cadmium telluride, these materials have relatively low toxicity and advantageous optical and electrical features, such as high absorption coefficients and tunable bandgaps [31]. Quaternary semiconductors like CZTS, CZTSe, and CZTSSe can all be produced inexpensively and sustainably [32]. Furthermore, they have the potential to provide reduced costs per watt of energy produced compared to conventional silicon-based solar cells. They are also suitable for usage in tandem solar cells, which can further boost their efficiency, due to their capacity to absorb a broad range of sunlight spectrum [33]. However, before CZTS, CZTSe, and CZTSSe can be used as replacement materials for solar cells, there are still several issues that need to be resolved. Increasing high conversion efficiencies, which are currently lower than those of silicon-based solar cells, is a major challenge [33]. Enhancing the reproducibility and stability of these materials, which can be impacted by elements such as the synthesis method, composition, and processing conditions, is another difficulty. There are various materials used in solar cell fabrication with Various narrow band gap semiconducting materials such as CIGS, $CuInS_2$, PbS, CdSe, CdTe, Bi_2S_3, Sb_2S_3. CZTS is the most promising but less studied compound as compared to other compounds for solar cell application. Moreover, recently advanced technology leads to the development of a variety of chalcogenide quantum dots such as binary metal chalcogenides QDs and alloyed QDs [34, 35]. The application of artificial intelligence, machine learning, and simulation can accelerate the screening and discovery of novel materials with potential properties for certain applications [36, 37]. The doping-inducted heterojunction is an effective strategy to boost the fabrication of semiconductor devices and photocatalysis [38, 39]. The heterojunction-based solar cells have become the front-runner among the next-generation thin-film photovoltaic technology due to their rapid great improvements in photovoltaic performance [40, 41]. The temperature effect of Solar Cells will affect the intrinsic properties of Solar Cell materials and the parameters that characterize Solar Cell performance. This will ultimately affect its power generation efficiency [42, 43].

It is known that the existence of defects in the absorber layer of solar cells is normally detereoriated the performance device. These defects such as cracks, poorly soldered joints, and mismatches occured in different layers lead to the high resistance and become hot spots in the thin film solar cells. Defects originated from interface of heterojunction also induce the deep energy levels in the semiconductor bandgap, which degrade the carrier lifetime and quantum efficiency of solar cells [44, 45]. Thus, it is necessary to improve the quality of materials, the fabrication process, and suitable doping on the solar cell to control the interfacial defects [46]. The origin of the defects is either introduced impurities or crystallographic defects such as vacancies or dislocations which are generated during the device fabrication. One has to reduce these defects at interfacial or bulk level which will not affect the photocurrent of the solar cells [47]. These defects in solar cell can be evaluated by several ways depending

on the nature and origin of the defects. Some common measurement tools such as dark current measurements, capacitance–voltage measurements, photoluminescence, electroluminescence, and I–V characteristics measurements have been used to examine the defects [48–50]. This paper also investigates the potential of Cu_2ZnSnS_4 (CZTS) as a solar absorber material for photovoltaic applications, given its eco-friendly, earth-abundant, and low-cost nature [32, 33]. Previous theoretical and experimental research on CZTS solar cells is also reviewed and CZTS was first used as a solar cell material in 1988 at Shinshu University in Japan. Researchers utilized sputtering to deposit thin films of CZTS, leading to the discovery of p-type conductive material, direct bandgap energy of 1.45 eV, and high absorption coefficients in the visible region [51]. While reported efficiencies have been much lower than the physical limit of 32% [52], Wang et al. demonstrated the highest power conversion efficiency (PCE) of 12.6%. Most simulation evaluations have been carried out using 1D simulation platforms like SCAPS or AMPS, which cannot account for the 3D properties of solar cells [53–55]. However, Lin et al. achieved an efficiency of 19.28% using a device structure based on CZTS as the absorber layer, Al-doped ZnO as a front electrode, un-doped ZnO as the window layer, and In_2S_3 as the buffer layer, with the help of the Solar Cell Capacitance Simulator (SCAPS) simulation tool [56]. The highest reported PCE of 25.30% was achieved by using a CZTS-based solar cell structure (Al:ZnO | i–ZnO/ZnSe/CZTS/Se) with an absorber layer thickness of 2 μm and a buffer layer thickness of 50 nm, with V_{OC} = 1.10 V, J_{SC} = 25.84 mA/cm^2, and Fill Factor (FF) = 88.47% [57]. These results demonstrate the feasibility of creating highly efficient CZTS solar cells. To increase the PCE, the Silvaco Atlas modeling program was used to create and simulate the CZTS solar cell, exploring its energy bandgaps, photogeneration and recombination rates, and the effect of defect density on PCE.

In this study, we explore the parameters for enhancing the absorber layer of CZTS-based solar cells by modifying the defect density. The validation of our findings is further confirmed through Silvaco Atlas simulation software. This paper presents an analysis of the photogeneration rate, recombination rate, and solar cell performance in response to variations in the absorber layer's defect density. Our results demonstrate that at a defect density of 1.8×10^{12} cm^{-3}, the solar cell achieves a peak efficiency of approximately 26.58%. These promising outcomes indicate the possibility of developing highly efficient CZTS solar cells.

2 Theoretical details

2.1 Simulation tool

The Simulation tool which has been applied in this scheme is Silvaco Atlas. It is applicable, particularly in simulating semiconductor devices. It foresees a device containing electrical properties and, at the same time, it can be observed in at least two dimensions: two dimensions (2D) and three dimensions (3D). The simulation uses several programs, including Deckbuild, Devedit, Tonyplot, and Athena, along with several subprograms that are accessible during the simulation and perform more specialized tasks that combine to produce realistic results [58]. Here, integrated programs facilitate the users in simulating the prototypes of semiconductor devices on their fundamental features. As a result, the programs exist with several prototypes, numerical systems, and many other types of inherent mechanisms which have been automated into the system, thereby giving a wide range of possibilities for operating the systems by its users [59].

2.2 Device structure

The proposed solar cell Al: ZnO/i–ZnO/CdS/CZTS/Mo is supposed to be a single-junction solar cell device model with CZTS as the absorber layer material. The front electrical contact made of Al: ZnO has been selected because it is transparent to incident solar radiation. The n-type CdS layer serves as the buffer layer and forms a hetero-structure p–n junction with the p-type CZTS absorber layer, while the i–ZnO layer has been used as the window layer and serves as a passivation layer. Table 1 lists the material parameters that were chosen for this study. These parameters were taken from published works [56, 60, 61]. Figure 1 depicts the device topology of the proposed CZTS-based solar cells.

Table 1 Calibrated CZTS structure of material parameters in solar cell

Parameters	ZnO:Al [60]	i–ZnO [73]	CdS [27]	CZTS [60]
Thickness (nm)	450	80	100	600
Relative permittivity (ε_r)	9	9	10	10
Band gap Eg (eV)	3.3	3.3	2.4	1.45
Electron affinity χ (eV)	4.4	4.4	4.2	4.1
Effective conduction band density N_c (cm^{-3})	2.2×10^{18}	2.2×10^{18}	2.2×10^{18}	2.2×10^{18}
Effective valence band density N_v (cm^{-3})	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}
Electron mobility μ_n (cm^2. V^{-1}. S^{-1})	100	100	100	5
Hole mobility μ_p (cm^2. V^{-1}. S^{-1})	25	25	25	1
Donor concentration N_d (cm^{-3})	1×10^{18}	–	1×10^{17}	–
Acceptor concentration Na (cm^{-3})	–	–	–	1×10^{16}
Gaussian defect density for the donor (cm^{-3})	–	–	–	1.8×10^{13}
Gaussian defect density for acceptor (cm^{-3})	1.8×10^{16}	–	1.8×10^{17}	–

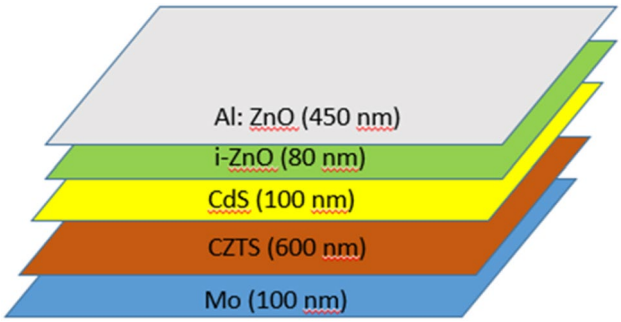


Fig. 1 Proposed device structure of CZTS-based solar cells

3 Result and discussion

3.1 Optimization of the thickness of the absorber layer

The performance of a solar cell is significantly affected by the thickness of its absorber layer. In this investigation, we examined the effect of changing the thickness of the CZTS absorber layer. It was discovered that CZTS demonstrates exceptional absorption in the blue region compared to other absorber materials, primarily due to its high absorption coefficient (> 104 cm^{-1}) [62]. The simulation encompassed a range of thicknesses for the solar cell, from 100 to 1000 nm, and the results are depicted in Fig. 2. The simulation results indicate that a thickness of 600 nm is optimal for achieving nearly complete absorption of AM 1.5G radiation, thanks to the material's strong absorption coefficient and direct bandgap [63]. As the CZTS layer increases in thickness, both the Jsc (short-circuit current density) and Voc (open-circuit voltage) of the solar cell also increase, resulting in improved efficiency and fill factor. The greater absorber layer absorbs more photons with longer wavelengths, facilitating the generation of electron–hole pairs, which accounts for this effect [64].

3.2 Optimization of defect density on the performance of the cell

The efficiency of the solar cell is greatly influenced by the defect density. The maximum PV parameter values, such as Voc, Jsc, FF, and, were observed up

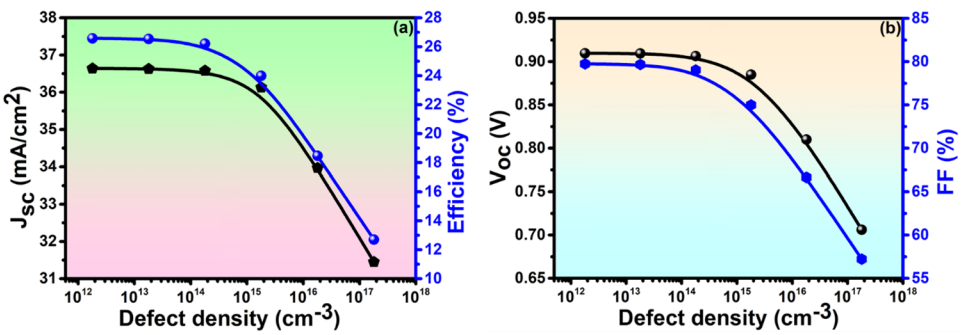


Fig. 2 Impact of defect density on **a** short-circuit current density and efficiency **b** Open-circuit voltage and fill factor

to a defect density of 10^{14} cm^{-3} [65]. This effect was observed in the light-absorbing CZTS layer. The flaw density of the absorber layer CZTS has been adjusted in this work from 1.8×10^{12} to 1.8×10^{17} cm^{-3}. While the fault density is less than 1.8×10^{14} cm^{-3}, it can be seen that the performance is consistent. The output factors are substantially impacted when the fault density exceeds 1.8×10^{14} cm^{-3}, though. Because there are more recombination centers available due to the rise in defect states in the absorber layer, higher recombination of generated carriers occurs there, which lowers Jsc, Voc, efficiency, and fill factor as illustrated in Fig. 3.

3.3 Impact of photoabsorption rate and energy bandgap

The number of photons absorbed per unit area of the solar cell per unit of time in terms of wavelength for a CZTS solar cell [66] is shown in Fig. 4a. The Photon absorption rate is higher at the junction between the CdS buffer layer and the CZTS absorber layer because more photons can be absorbed due to less energy bandgap of the CZTS layer, after that, there is a sharp decrease. Almost 97% percent of a photon can be absorbed with the 600 nm thick CZTS layer, it is because photons having energy higher or equal to 1.46 eV are absorbed in the CZTS layer [67].

The bandgap energy is the minimum energy that is needed to knock out electrons from the lower energy state to the conduction energy state where they can contribute to photogeneration [68]. The energy band diagram of the CZTS solar cell is shown in Fig. 4b. The energy bandgap of ZnO is higher (3.37 eV), as a result, the lower wavelength of the light is absorbed by the ZnO in all other wavelengths and all the bottom CdS buffer layers are absorbed mid-range or visible region spectrum, and where the CZTS absorber layer is used to absorb higher wavelength. The CdS buffer and CZTS absorber layer are used to control the recombination and form a heterojunction with the CZTS layer [60].

3.4 Impact of photogeneration and recombination rate

One of the key features of the solar cell that improves its overall performance is the photogeneration and recombination process. The total number of electrons produced in the PV cell as a result of light absorption is expressed as the photogeneration rate [67]. The photogeneration rate is almost constant at 21 photons

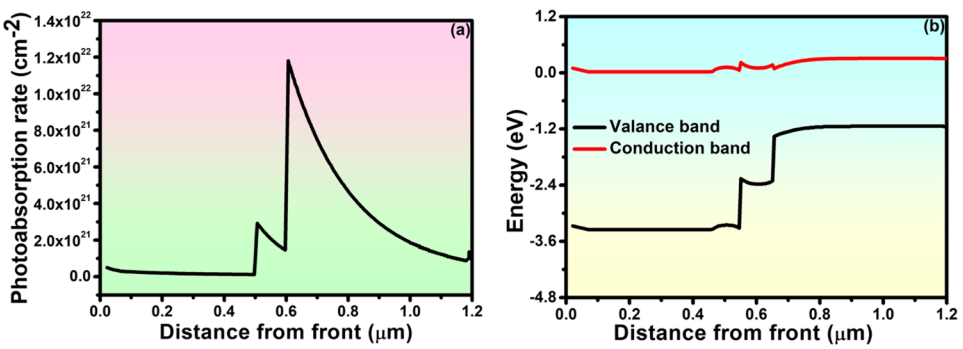


Fig. 3 Analysis of **a** photoabsorption rate and **b** Band gap energy with distance from front

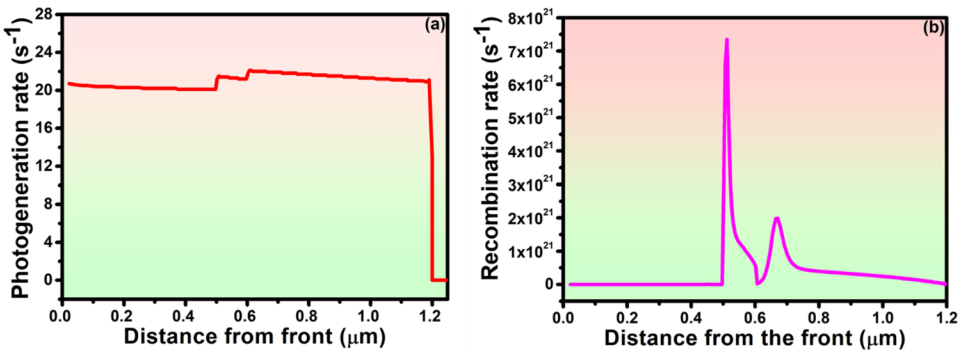


Fig. 4 Analysis of **a** photogeneration rate and **b** Recombination rate with distance from front

per cm^3 which is mostly due to radiative recombination, and which restricts the efficiency of the solar cell. In a CZTS-based cell probability of radiative recombination increases that of non-radiative recombination when there is a direct transition of the carrier between the conduction and valence band [69]. The photogeneration rate of the purposed CZTS solar cell is shown in Fig. 5a. Similarly, the recombination rate is higher at the junction between the CdS buffer layer and the CZTS absorber layer as described in Fig. 5b. The sunlight produced minority carriers of electrons and has more probability to recombine with a hole in the absorber and increasing the recombination rate can decrease efficiency as well. The decrease in crystallinity in the interface between the CdS and CZTS layer leads to a reduction in the efficiency of CZTS-based solar cells [60].

3.5 Analysis of electric potential and electric field

In the mechanism for energy loss within the solar cell, the distribution of the electrical field is crucial [70]. Electrostatic potential fluctuations suffer because CZTSSe has a lower dielectric constant than CIGS [71]. Within the CZTS absorber layer, the electric potential is larger than outside, and it then begins to decrease as seen in Fig. 6a. To improve the transport mechanism of the photogenerated carriers, the CZTS absorber layer electric field can be adjusted by altering the Se content in the absorber layer [72]. Electric field spikes can be visible at the CdS buffer layer and the CZTS absorber layer in Fig. 6b. The buffer layer is essential to the creation of an electric field at the junction in the space charge zone [73].

3.6 Photovoltaic (PV) properties

The performance of solar cells depends upon Isc, Voc, FF, and efficiency. After the simulation of the solar cell using the parameters shown above in Table 1 by the Silvaco atlas, I–V characteristics of the CZTS solar cell as shown in Fig. 7a have been obtained. It is seen that the PCE of solar cells has been observed at 26.58% with a defect density of 1.8×10^{12} and the Isc, Voc, and Fill factors have been observed as 36.64 mA/cm^2, 0.909 V, and 79.25% respectively. The reported PV properties of CZTS solar cells have been summarized in Table 2.

Quantum efficiency is the ratio of incident photons of particular energy to the number of carriers the solar collectors can capture [74]. As the wavelength of the solar cell increases from 300 nm, the quantum

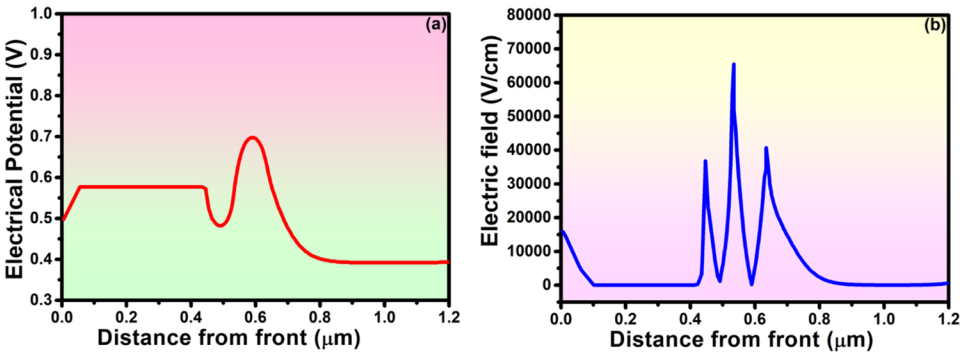


Fig. 5 Analysis of **a** electrical potential and **b** electric field with distance from front

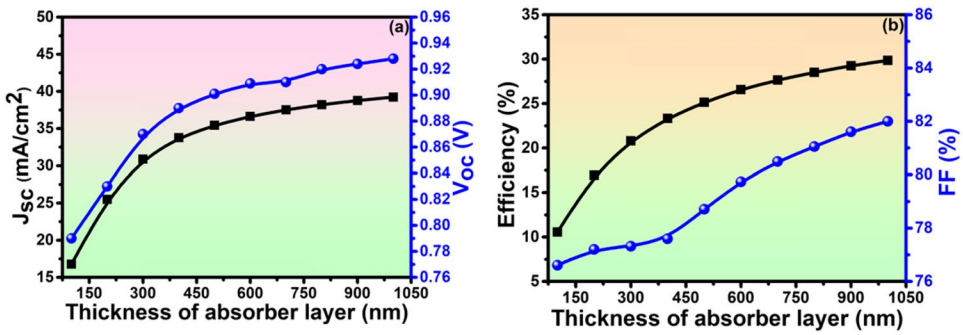


Fig. 6 Analysis of **a** Jsc &Voc and **b** Efficiency & FF with the thickness of absorber layer CZTS-based solar cell

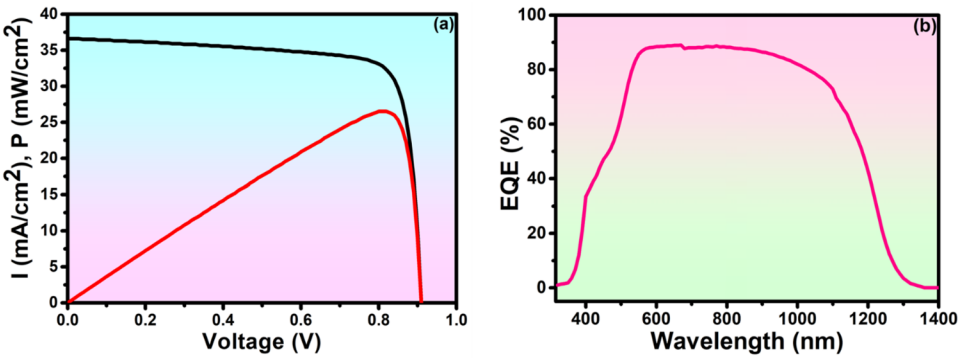


Fig. 7 Analysis of **a** I, P–V characteristics and **b** External quantum efficiency of CZTS-based solar cell

Table 2 Comparative study of reported PV properties of solar cells

Structure of solar cell	Isc (mA/cm^2)	Voc (V)	FF (%)	Efficiency (%)	References
i–ZnO/n–CdS/p–CZTS/Mo	23.51	0.840	76.28	15.04	[75]
ZnO:Al/i–ZnO/CdS/CZTS/Mo	19.50	0.748	57.48	8.40	[76]
ITO/TiO_2/CZTZ	30.99	0.920	77.38	22.15	[77]
Al:ZnO|iZnO|ZnSe|CZTS|Se)	25.84	1.100	88.47	25.30	[57]
CZTSe QDs	13.65	0.470	56.00	3.61	[78]
CZTSe QDs/TiO_2 NFs	11.28	0.45	55.00	2.84	[78]
CdS/$CuInS_2$	28.28	0.57	78.97	12.74	[79]
I–ZnO/n–CdS/p–CZTS/Mo	36.64	0.909	79.74	26.58	This work

efficiency increases. At the wavelength between 550 and 900 nm quantum efficiency is almost constant at some value and then starts decreasing rapidly from 100 nm wavelength as shown in Fig. 7b. It is noted that the external quantum efficiency (EQE) reached nearly 90% at wavelengths 550 to 900 nm.

4 Conclusions

ZnO–Al/i–ZnO/n–CdS/CZTS/Mo combination was subjected to numerical analysis and optimization using Silvaco Atlas software. The study was aimed to investigate the influence of layer thickness and doping density on the solar cell efficiency. The simulation results demonstrate that the solar cell can achieve a maximum efficiency of approximately 26.58% when the defect density is at 1.8 × 1012 cm^{-3}, accompanied by Isc = 36.64 A and Voc = 0.909 V. Furthermore, an increase in layer thickness has been observed to enhance the efficiency of the solar cell due to a greater absorption of photons. The effect of defect density on the absorber layer, photogeneration, and recombination processes on solar cell efficiency has also been examined. By optimizing the cell parameters, an improved power conversion efficiency (PCE) of 26.58% and a fill factor (FF) of 79.74% were achieved, underscoring the potential of CZTS-based solar cells as cost-effective and highly efficient alternatives.

Acknowledgements

This work was supported by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No.20204010600470). This research was supported by research funds of Jeonbuk National University in 2023. The authors are thankful to the Deanship of Scientific Research at Najran University, Najran, Kingdom of Saudi Arabia for funding under the Research Group funding program Grant No. NU/RG/SERC/12/49.

Author contributions

This work is the collaborative development of all the authors. DKC: original draft, software, conceptualization. DKS: editing-draft, methodology, revision. SK: investigation, formal analysis. NB: data

curation, investigation, validation. DRA: data curation, resources. KBK: software. AU: investigation, data curation, editing-draft, writing-review, and editing. AI, MAMA: data curation. MSA, SB: writing-review and editing. OBA: funding acquisition, writing review, supervision, and editing. All authors have read and agreed to the published version of the manuscript.

Funding

The authors have not disclosed any funding.

Data availability

All data generated or analyzed during this study are included in this published article.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. A. Umar, P.K. Singh, D.K. Dwivedi, A.A. Ibrahim, M.A.M. Alhamami, H. Qasem, S. Akbar, S. Baskoutas, Design and simulation of lead-free perovskite solar cells with a hole transport layer made of NiO nanocomposite. Sci. Adv. Mater. **14**, 1511–1517 (2022)
2. D.K. Shah, K.C. Devendra, D. Parajuli, M.S. Akhtar, C.Y. Kim, O.-B. Yang, A computational study of carrier lifetime, doping concentration, and thickness of window layer for GaAs solar cell based on Al_2O_3 antireflection layer. Sol. Energy **234**, 330–337 (2022)
3. J. Wang, R. Fu, S. Wen et al., Progress and current challenges for CO_2 capture materials from ambient air. Adv. Compos. Hybrid Mater. **5**, 2721–2759 (2022)
4. A. Muhammad, S.U. Rather, M. Umar, H. Bamufleh, A.M. Ali, T.E. Youssef, Carbon dioxide (CO_2) capture in alkanolamines impregnated activated carbon developed from date stones. Sci. Adv. Mater. **13**, 98–104 (2021)
5. E. Chamanehpour, M.H. Sayadi, M. Hajiani, A hierarchical graphitic carbon nitride supported by metal–organic framework and copper nanocomposite as a novel bifunctional catalyst with long-term stability for enhanced carbon dioxide photoreduction under solar light irradiation. Adv. Compos. Hybrid Mater. **5**, 2461–2477 (2022)
6. Y. Li, L. Li, S. Luo et al., The role of K in tuning oxidative dehydrogenation of ethane with CO_2 to be selective toward ethylene. Adv. Compos. Hybrid Mater. **4**, 793–805 (2021)
7. D. Ping, F. Yi, G. Zhang, S. Wu, S. Fang, K. Hu, B.B. Xu, J. Ren, Z. Guo, NH_4Cl-assisted preparation of single Ni sites anchored carbon nanosheet catalysts for highly efficient carbon dioxide electroreduction. J. Mater. Sci. Technol. **142**, 1–9 (2023)
8. D.K. Shah, K.C. Devendra, M. Muddassir, M.S. Akhtar, C.Y. Kim, O.-B. Yang, A simulation approach for investigating the performances of cadmium telluride solar cells using doping concentrations, carrier lifetimes, thickness of layers, and band gaps. Sol. Energy **216**, 259–265 (2021)
9. A. Shaheen, S. Hussain, G.J. Qiao, M.H. Mahmoud, H. Fouad, M.S. Akhtar, Outstanding electrochemical supercapacitor performances of $NiCo_2O_4$ nanoflowers. Sci. Adv. Mater. **13**, 2460–2466 (2021)
10. C. Lai, Y. Guo, H. Zhao et al., High-performance double "ion-buffering reservoirs" of asymmetric supercapacitors enabled by battery-type hierarchical porous sandwich-like Co_3O_4 and 3D graphene aerogels. Adv. Compos. Hybrid Mater. **5**, 2557–2574 (2022)
11. W. Yang, D. Peng, H. Kimura et al., Honeycomb-like nitrogen-doped porous carbon decorated with Co_3O_4 nanoparticles for superior electrochemical performance pseudo-capacitive lithium storage and supercapacitors. Adv. Compos. Hybrid Mater. **5**, 3146–3157 (2022)
12. Y. Zhao, Y. Li, D. Zhang, S. Song, J. Wang, Y. Ke, In-Situ grown sheet-like nanostructure of $NiCo_2S_4$ for high performance supercapacitors. Sci. Adv. Mater. **13**, 1065–1069 (2021)
13. K. Cui, L. Zhu, W. Guo, P. Shuang, X. Yang, X. Chai, W. Du, Recent advance of Ni–Co–X (X = O, S, Se, Te) bimetallic compound nanoarray electrode materials applied in supercapacitors. Sci. Adv. Mater. **14**(5), 819–828 (2022)
14. J. Yang, L. Tong, A.S. Alsubaie et al., Hybrid proton exchange membrane used in fuel cell with amino-functionalized metal–organic framework in sulfonated polyimide to construct efficient ion transport channel. Adv. Compos. Hybrid Mater. **5**, 834–842 (2022)
15. R.D. Prasad, C.B. Desai, O.P. Srivastava, S.R. Prasad, T.S. Bhat, B. Kamble, P. Sarvalkar, A. Kanthe, R.D. Kale, S. Banga, A. Patil, A. Saxena, S. Saxena, K. Saxena, A.K. Sharma, R.Y. Prasad, A critical review on recent developments in advanced supercapacitors for veterinary medicine. ES Food Agrofor. **11**, 805 (2023)

16. Mu. Wei Pan, D.X. Zhang, X. Sun, Supercapacitance property study of 3D open-framework prussian blue in neutral electrolyte. Sci. Adv. Mater. **13**, 438–448 (2021)
17. C. Gao, W. Deng, F. Pan, X. Feng, Y. Li, Superhydrophobic electrospun PVDF membranes with silanization and fluorosilanization co-functionalized CNTs for improved direct contact membrane distillation. Eng. Sci. **9**, 35–43 (2020)
18. F. Ahmed, P.M.Z. Hasan, S. Kumar, N.M. Shaalan, A. Aljaafari, N. Arshi, M. Albossed, G. Almutairi, B. Alotaibi, A facile method for the preparation of a-Fe_2O_3/reduced graphene oxides nanocomposites as electrode materials for high performance supercapacitors. Sci. Adv. Mater. **14**, 1342–1347 (2022)
19. C. Dang, Q. Mu, X. Xie et al., Recent progress in cathode catalyst for nonaqueous lithium oxygen batteries: a review. Adv. Compos. Hybrid Mater. **5**, 606–626 (2022)
20. D.K. Maurya, R. Dhanusuraman, Z. Guo et al., Composite polymer electrolytes: progress, challenges, and future outlook for sodium-ion batteries. Adv. Compos. Hybrid Mater. **5**, 2651–2674 (2022)
21. H. Soonmin, S. Wagh, A. Kadier, I.A. Gondal, N.P.B.A. Azim, M.K. Mishra, Renewable energy technologies. Sustain. Innov. Impact **2018**, 237–250 (2018)
22. D. Kc, R. Wagle, R. Gaib, A. Shrivastava, L.N. Mishra, Modelling and simulation of AlGaAs/GaAs solar cell. Am. J. Eng. Res. **9**, 4 (2020)
23. D.K. Shah, K.C. Devendra, T.-G. Kim, M.S. Akhtar, C.Y. Kim, O.-B. Yang, Influence of minority charge carrier lifetime and concentration on crystalline silicon solar cells based on double antireflection coating: a simulation study. Opt. Mater. **121**, 111500 (2021)
24. J. Müller, B. Rech, J. Springer, M. Vanecek, TCO and light trapping in silicon thin film solar cells. Sol. Energy. **77**, 917–930 (2004)
25. A. Polman, M. Knight, E.C. Garnett, B. Ehrler, W.C. Sinke, Photovoltaic materials: present efficiencies and future challenges. Science **352**, 4424 (2016)
26. D.K. Shah, K.C. Devendra, J. Choi, S.H. Kang, M.S. Akhtar, C.Y. Kim, O.-B. Yang, Determinantal study on the thickness of graphene oxide as arc layer for silicon solar cells using: a simulation approach. Mater. Sci. Semicond. Process. **147**, 106695 (2022)
27. N. Asim et al., A review on the role of materials science in solar cells. Renew. Sustain. Energy Rev. **16**, 5834–5847 (2012)
28. P. Jackson et al., Effects of heavy alkali elements in Cu(In, Ga)Se_2 solar cells with efficiencies up to 22.6%. Phys. Status Solidi Rapid Res. Lett. **10**, 583–586 (2016)
29. X. Li, P. Li, Z. Wu, D. Luo, H.-Y. Yu, L. Zheng-Hong, Review and perspective of materials for flexible solar cells. Mater. Rep.: Energy **1**(1), 10000 (2021)
30. J.D. Poplawsky et al., Structural and compositional dependence of the $CdTe_xSe_{1-x}$ alloy layer photoactivity in CdTe-based solar cells. Nat. Commun. **7**, 1–10 (2016)
31. K.J. Yang, D.H. Son, S.J. Sung, J.H. Sim, Y.I. Kim, S.N. Park, D.H. Kim, A band-gap-graded CZTSSe solar cell with 12.3% efficiency. J. Mater. Chem. A **4**(26), 10151–10158 (2016)
32. H. Katagiri, K. Jimbo, W.S. Maw, K. Oishi, M. Yamazaki, H. Araki, A. Takeuchi, Development of CZTS-based thin film solar cells. Thin Solid Films **517**(7), 2455–2460 (2009)
33. C. Wadia, A.P. Alivisatos, D.M. Kammen, Materials availability expands the opportunity for large-scale photovoltaics deployment. Environ. Sci. Technol. **43**, 2072 (2009)
34. A.T. Supekar, P.K. Bhujbal, S.A. Salunke, S.M. Rathod, S.P. Patole, H.M. Pathan, Bismuth sulfide and antimony sulfide-based solar cells: a review. ES Energy Environ. **19**, 848 (2023)
35. A. Umar, P.K. Singh, D.K. Dwivedi, A.A. Ibrahim, M.A.M. Alhamami, S. Baskoutas, Design and simulation of lead-free perovskite solar cells with a hole transport layer made of NiO nanocomposite. Sci. Adv. Mater. **14**, 1539–1545 (2022)
36. A. Umar, P. Srivastava, S. Rai, P. Lohia, D.K. Dwivedi, H. Algadi, S. Baskoutas, High-performance lead-free perovskite solar cell: a theoretical study, emerging. Mater. Res. **12**, 1–6 (2023)
37. Y. Cao, M. Weng, M.H.H. Mahmoud et al., Flame-retardant and leakage-proof phase change composites based on MXene/polyimide aerogels toward solar thermal energy harvesting. Adv. Compos. Hybrid Mater. **5**, 1253–1267 (2022)
38. E.-B. Kim, M.S. Akhtar, S. Ameen, A. Umar, H. Qasem, H.-G. Rubahn, M. Shkir, A. Kaushik, Y.K. Mishra, Improving the performance of 2D perovskite solar cells by carrier trappings and minifying the grain boundaries. Nano Energy **102**, 107673 (2022)
39. Y. Guo, T. Liu, H. He et al., Bifunctional interface modification for efficient and UV-robust α-Fe_2O_3-based planar organic–inorganic hybrid perovskite solar cells. Adv. Compos. Hybrid Mater. **5**, 3212–3222 (2022)
40. P. Srivastava, S. Rai, P. Lohia, D.K. Dwivedi, H. Qasem, A. Umar, S. Akbar, H. Algadi, S. Baskoutas, Theoretical study of perovskite solar cell for enhancement of device performance using SCAPS-1D. Phys. Scripta **97**, 125004 (2022)

41. H. Gao, J. Li, Y. Liu et al., Shape memory polymer solar cells with active deformation. Adv. Compos. Hybrid Mater. **4**, 957–965 (2021)
42. A. Umar, P.K. Singh, D.K. Dwivedi, H. Algadi, A.A. Ibrahim, M.A.M. Alhamami, S. Baskoutas, High power-conversion efficiency of lead-free perovskite solar cells: a theoretical investigation. Micromachines **13**(12), 2201 (2022)
43. C. Sun, Y. Zou, C. Qin et al., Temperature effect of photovoltaic cells: a review. Adv. Compos. Hybrid Mater. **5**, 2675–2699 (2022)
44. H. Ullah, B. Marí, S. Ruiz, Effect of defects on the performance of some photovoltaic solar cells: an introduction to research methods to engineering students. SEFI Conf. **1**, 12–15 (2016)
45. O. Breitenstein, J. Bauer, P.P. Altermatt, K. Ramspeck, Influence of defects on solar cell characteristics. Solid State Phenom. **156**, 1–10 (2010)
46. M.A. El-Rashidy, An efficient and portable solar cell defect detection system. Neural Comput. Appl. **34**, 18497–18509 (2022)
47. A. Zekry, A. Shaker, M. Salem, Solar cells and arrays: principles, analysis, and design, in *Advances in renewable energies and power technologies*. (Elsevier, Amsterdam, 2018), pp.3–56
48. J. R. Davis, Capacitance measurements of defects in solar cells: checking the model assumptions. (2015).
49. S. Binetti, M. Acciarri, J. Libal, Impact of extended defects on the electrical properties of solar grade multicrystalline silicon for solar cell application. Solid State Phenom. **131**, 419–424 (2008)
50. K. Buehler, K. Kaufmann, M. Patzold, M. Sprenger, S. Schoenfelder, Identifying defects on solar cells using magnetic field measurements and artificial intelligence trained by a finite-element-model. EPJ Photovolt. **14**, 12 (2023)
51. Y.E. Romanyuk, C.M. Fella, A.R. Uhl, M. Werner, A.N. Tiwari, T. Schnabel, E. Ahlswede, Recent trends in direct solution coating of kesterite absorber layers in solar cells. Sol. Energy Mater. Sol. Cells. **119**, 181 (2013)
52. K. Ito, T. Nakazawa, Electrical and optical properties of stannite-type quater-nary semiconductor thin films. Jpn. J. Appl. Phys. **11**, 2094–2097 (1988)
53. W. Wang, M.T. Winkler, O. Gunawan, T. Gokmen, T.K. Todorov, Y. Zhu, D.B. Mitzi, Device characteristics of CZTSSe thin-film solar cells with 12.6% efficiency. Adv. Energy Mater. **4**(7), 1301465 (2014)
54. A. Cherouana, R. Labbani, Study of CZTS and CZTSSe solar cells for buffer layers selection. Appl. Surf. Sci. **424**(2), 251–255 (2017)
55. W. Zhao, W. Zhou, X. Miao, Numerical simulation of CZTS thin film solar cell, NEMS proceedings of IEEE, Kyoto, Japan, March, 502–506 (2012).
56. P. Lin, L. Lin, J. Yu, S. Cheng, P. Lu, Q. Zheng, Numerical simulation of Cu_2ZnSnS_4 based solar cells with In_2S_3 buffer layers by SCAPS-1D. J. Appl. Sci. Eng. **17**(4), 383–390 (2014)
57. D. Talukder, R. Pal, M.N. Hasan, M.F. Wahid, Numerical simulation of CZTS solar cell with ZnSe buffer layer. In 2021 International conference on automation, control and mechatronics for industry 4.0 (ACMI) (1–6) IEEE (July 2021).
58. Atlas User's Manual, Silvaco International, Santa Clara, CA, (2005).
59. Atlas User's Manual, Silvaco International, Santa Clara, CA, (2016).
60. X. Liu, X. Hao, S. Huang, G. Conibeer, Numerical modeling of CZTS solar cells, 39th IEEE photovoltaic specialists conference (2013).
61. S. Fadili, B. Hartiti, A. Kotbi, A. Ridah, P. Thevenin, Numerical simulation of solar cells based CZTS buffer layer ($ZnO_{1-X}S_X$) using SCAPS-1D software. J. Fund. Appl. Sci. **9**(2), 1001–1011 (2017)
62. R. Tiwari, P. Dubey, P. Lohia, D.K. Dwivedi, H. Fouad, M.S. Akhtar, Simulation engineering in quantum dots for efficient photovoltaic solar cell using copper iodide as hole transport layer. J. Nanoelectron. Optoelectron. **16**(12), 1897–1904 (2021)
63. K. Zhang, Z. Su, L. Zhao, C. Yan, F. Liu, H. Cui, X. Hao, Y. Liu, Improving the conversion efficiency of Cu_2ZnSnS_4 solar cell by low pressure sulfurization. Appl. Phys. Lett. **104**, 141101 (2014)
64. O.K. Simya, A. Mahaboobbatcha, K. Balachande, A comparative study on the performance of Kesterite based thin film solar cells using SCAPS simulation program. Superlattices Microstruct. **82**, 248–261 (2015)
65. A.S. Mathur, S. Upadhyay, P.P. Singh, B. Sharma, P. Arora, V.K. Rajput, B.P. Singh, Role of defect density in absorber layer of ternary chalcogenide Cu_2SnS_3 solar cell. Opt. Mater. **119**, 111314 (2021)
66. M. Mirzaei, J. Hasanzadeh, A.A. Ziabari, Efficiency enhancement of CZTS solar cells using Al plasmonic nanoparticles: the effect of size and period of nanoparticles. J. Electron. Mater. **49**(12), 7168–7178 (2020)
67. http://www.pveducation.org/pvcdrom/pn-junction/band-gap (Accessed on 21 March 2023)
68. D. Kc, D.K. Shah, A. Shrivastava, Computational study on the performance of zinc selenide as window layer for efficient GaAs solar cell. Mater. Today: Proc. **49**, 2580–2583 (2022)

69. K. Ito, *Copper zinc tin sulfide-based thin-film solar cells* (John Wiley & Sons, 2014)
70. M. Cai, N. Ishida, X. Li, X. Yang, T. Noda, Y. Wu, L. Han, Control of electrical, potential distribution for high-performance perovskite solar cells. Joule **2**(2), 296–306 (2018)
71. T.J. Huang, X. Yin, G. Qi, H. Gong, CZTS-based materials and interfaces and their effects on the performance of thin film solar cells. Phys. Status Solidi Rapid Res. Lett. **8**(09), 735–762 (2014)
72. H. Ferhati, F. Djeffal, Graded band-gap engineering for increased efficiency in CZTS solar cells. Opt. Mater. **76**, 393–399 (2018)
73. F. Belarbi, W. Rahal, D. Rached, M. Adnane, A comparative study of different buffer layers for CZTS solar cell using scaps-1D simulation program. Optik **216**, 164743 (2020)
74. R. Mahbub, M. Islam, S. Anwar, F. Satter, S.S. Satter, S.M. Ullah, Simulation of CZTS thin film solar cell for different buffer layers for high efficiency performance. South Asian J. Eng. Technol. **2**(52), 1–10 (2016)
75. A. Benami, Effect of CZTS parameters on photovoltaic solar cell from numerical simulation. J. Energy Power Eng. **13**, 32–36 (2019)
76. M.D. Wanda, S. Ouédraogo, F. Tchoffo, F. Zougmoré, J.M.B. Ndjaka, Numerical investigations and analysis of Cu_2ZnSnS_4 based solar cells by SCAPS-1D. Int. J. Photoenergy (2016). https://doi.org/10.1155/2016/2152018
77. M.A. Rahman, Enhancing the photovoltaic performance of Cd-free Cu_2ZnSnS_4 heterojunction solar cells using SnS HTL and TiO_2 ETL. Sol. Energy **215**, 64–76 (2021)
78. R. Kottayi, D.K. Maurya, R. Sittaramane, S. Angaiah, Recent developments in metal chalcogenides based quantum dot sensitized solar cells. ES Energy Environ. **18**, 1–40 (2022)
79. B.M. Sakunde, S. Patole, H.M. Pathan, Numerical modeling to improve the efficiency of cadmium sulfide/copper indium sulfide ($CdS/CuInS_2$) thin film-based solar cells. ES Energy Environ. **18**, 111–121 (2022)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

View publication stats

algorithms

MDPI

Article

Ensemble Transfer Learning for Distinguishing Cognitively Normal and Mild Cognitive Impairment Patients Using MRI

Pratham Grover [1], Kunal Chaturvedi [2], Xing Zi [2], Amit Saxena [3], Shiv Prakash [4], Tony Jan [5] and Mukesh Prasad [2,*]

[1] Department of Biotechnology, Delhi Technological University, Bawana Road, Delhi 110042, India
[2] School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney 2007, Australia
[3] Department of Computer Science and Information Technology, Guru Ghasidas University, Bilaspur 495009, India
[4] Department of Electronics and Communication, University of Allahabad, Allahabad 211002, India
[5] School of Information Technology, Torrens University, Sydney 2010, Australia
* Correspondence: mukesh.prasad@uts.edu.au

Abstract: Alzheimer's disease is a chronic neurodegenerative disease that causes brain cells to degenerate, resulting in decreased physical and mental abilities and, in severe cases, permanent memory loss. It is considered as the most common and fatal form of dementia. Although mild cognitive impairment (MCI) precedes Alzheimer's disease (AD), it does not necessarily show the obvious symptoms of AD. As a result, it becomes challenging to distinguish between mild cognitive impairment and cognitively normal. In this paper, we propose an ensemble of deep learners based on convolutional neural networks for the early diagnosis of Alzheimer's disease. The proposed approach utilises simple averaging ensemble and weighted averaging ensemble methods. The ensemble-based transfer learning model demonstrates enhanced generalization and performance for AD diagnosis compared to traditional transfer learning methods. Extensive experiments on the OASIS-3 dataset validate the effectiveness of the proposed model, showcasing its superiority over state-of-the-art transfer learning approaches in terms of accuracy, robustness, and efficiency.

Keywords: Alzheimer's disease; ensemble learning; classification; convolutional neural network; mild cognitive impairment

check for updates

Citation: Grover, P.; Chaturvedi, K.; Zi, X.; Saxena, A.; Prakash, S.; Jan, T.; Prasad, M. Ensemble Transfer Learning for Distinguishing Cognitively Normal and Mild Cognitive Impairment Patients Using MRI. *Algorithms* **2023**, *16*, 377. https://doi.org/10.3390/a16080377

Academic Editor: Frank Werner

Received: 13 June 2023
Revised: 31 July 2023
Accepted: 3 August 2023
Published: 6 August 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Alzheimer's disease (AD), the most common form of dementia, poses a significant challenge to healthcare providers. It causes a progressive, irreversible brain disorder characterised by a deterioration of cognitive functions. In the United States, Alzheimer's disease is the fifth leading cause of death among adults age 65 and older, and there was a 145% increase in deaths from Alzheimer's disease in 2017 [1]. Recently, a significant amount of research [2–6] has been conducted in order to detect early and pre-symptomatic stages of AD to slow or prevent the progression of the disease. Surprisingly, simple cognitive assessments of dementia were only accepted by a minority of the older adults. In particular, techniques based on MRI testing have been developed and used to identify AD [5]. The accuracy of an AD diagnosis mostly depends on the biomarkers of the disease. The features to detect Alzheimer are extracted from the region of interest (ROI) of the brain which has the potential to contain the tissue that are sensitive markers for dementia. Similarly, the distinctive features are confirmed by the histopathology of dementia that display these ROI in the amygdala and hippocampus embedded deep into the temporal lobe of the brain [7,8]. Due to the recent advancements in automating the prediction of early diagnosis of dementia, the objective has shifted to help non-experts and non-neurologist users to identify early signs of AD.

AI-based computer-aided diagnosis (CAD) has become a feasible and a popular tool in medicine due to its low cost and transparent decision making [9–12]. Traditional methods of feature extraction require handcrafted features [13–16]. Castellazzi et al. [17] combined advanced qMRI measurements along with conventional machine learning techniques to classify AD from Vascular Dementia (VD) and investigated if the suggested method might be used to forecast the common illness in those who do not have distinct AD or VD profiles. In the work proposed by Battineni et al. [18], four machine learning algorithms were trained with three different sets of experiments—extracting features manually, extracting features automatically, and ensembling all the models. Alickovic et al. [19] manually extracted brain characteristics using the histogram approach in order to feed the data into the Random Forest model. The proposed method is not only a conventional process, but it is also time-consuming and leads to accuracy that is subpar.

On the other hand, deep learning (DL) approaches are incredibly effective since they automatically extract key characteristics from the data. Transfer learning may recognise new things in accordance with the knowledge that is learnt from other things, just like human cognition can. Based on the notion that transfer learning can efficiently use the knowledge of old domains to identify the goals of a new domain, numerous enhanced transfer learning methodologies have been presented in recent years. Ahana et al. [20] employed two strategies: one is ensembling different machine learning algorithms via two methods, gradient boosting and a voting classifier, and the second is training an artificial neural network (ANN) to detect AD. A lot of researchers have confined their studies of classification of AD to the ADNI dataset, as investigated in some papers [21–23]. However, ensembles of deep convolutional architectures and similar networks that use deep learning have produced the best results. Wang et al. [21] presented an ensemble of 3D densely connected convolutional networks (3D-DenseNets) for AD and MCI diagnosis. The use of a probabilistic-based ensemble method showed an increase in accuracy.

In this paper, we propose to use the ensemble technique by combining the predictive capabilities of various state-of-the-art pre-trained transfer learning models for AD diagnosis. At the beginning, pre-processed MRI images are passed on to each individual pre-trained model such as VGG19, DenseNet201, EfficientNetV2S, MobileNet, ResNet152, InceptionV3, NASNetLarge, and Xception. This method has been proven useful for the insufficient availability of large image datasets [24,25]. The following are our key contributions to this work:

- This paper presents a neural network framework with two ensemble methods, i.e., weighted averaging and simple averaging, on the OASIS-3 dataset for AD diagnosis.
- Fine tuning of eight different models such as VGG19, DenseNet201, EfficientNetV2S, MobileNet, ResNet152, InceptionV3, NASNetLarge, and Xception achieved a higher accuracy in comparison to the state of the art.
- This paper conducts a qualitative evaluation in the result variations of pre-trained models between cropping and without cropping of MRI images.

The rest of the research paper is structured as follows: In Section 2, related work is discussed. Section 3 entails the details of the workflow of the proposed framework along with the detailed discussion of the dataset and ensemble learning. Section 4 lays down the experiments and results performed on the provided dataset. Finally, Section 5 summarizes this study with future insights.

2. Related Works

The field of Alzheimer's disease (AD) classification has indeed made significant advancements with the utilisation of multimodal brain imaging data and deep learning approaches. Deep convolutional neural networks (CNNs) have emerged as powerful tools for AD classification due to their ability to leverage the rich information provided by various imaging modalities. Several research studies have been conducted to improve the classification of AD using deep learning techniques. Some studies focus on manually building feature extraction approaches combined with machine learning algorithms, while

others utilise deep learning methods. For example, a hybrid deep learning framework was introduced for the early-stage diagnosis and detection of AD [26], and improved CNNs were proposed for classifying AD using multimodal brain imaging data [27].

Liu et al. [28] presented a multi-modality cascaded convolutional network for the prediction and detection of AD. A systematic study on deep-learning-based medical image analysis for disease diagnosis was conducted by Abdou [29]. Additionally, a multi-scale feature fusion network was introduced for classifying illnesses like AD using MRI data [30], and a three-channel phase feature learning model was proposed for jointly learning from MRI and PET images for AD diagnosis [31]. Various deep learning methods, including CNNs, stacked deep polynomial networks, and generative adversarial networks (GANs), have also been explored for the diagnosis of AD. For instance, a tensor-train, high-order pooling and semisupervised learning-based GAN (THS-GAN) was presented by Yu et al. [32], and Shi et al. [33] developed multimodal deep polynomial networks.

CNNs can be used as feature extractors and classifiers, or solely as feature extractors. Some researchers use CNN to extract features and then apply traditional machine learning methods for classification [34]. CNN has been employed to extract target-level representations generated from sparse regression for clinical decision making [35]. Furthermore, a combination of 3D CNN with MRI was utilised for AD classification [36]. In specific studies, 2D CNNs have been developed and trained using 2D MRI slices as input. Wang et al. [37] incorporated a pre-trained VGG-16 feature extractor from ImageNet and explored the application of Lasso and PCA in conjunction with CNN to predict the conversion of mild cognitive impairment (MCI) to AD. Similarly, in [38], a landmark-based deep multi-instance learning (LDMIL) classifier was proposed, which can detect anatomical landmarks in brain MRI images and define regions of interest (ROIs) for diagnosis using the ADNI dataset.

Other approaches have focused on structural magnetic resonance imaging (sMRI) to improve AD diagnosis. For example, a dual-attention multi-instance deep learning model (DA-MIDL) was developed to extract global and local features and create a classification system [39]. Automatic segmentation methods using the ensembling of neural networks on the hippocampus region have also been employed [40,41]. Furthermore, a semi-supervised multimodal Laplacian regularized least squares (mLapRLS) method was developed to differentiate MCI subjects from healthy controls (HC) [42], and an architecture capable of combining multimodal neuroimaging information in one setting was proposed to classify AD [43]. Lastly, Liu et al. [44] developed a Monte Carlo ensemble neural network on 2D image slices from multiple database sources and achieved an accuracy of 90% for AD classification.

The field of Alzheimer's disease (AD) classification has made significant advancements using multimodal brain imaging data and deep learning approaches, particularly deep convolutional neural networks (CNNs). Deep learning eliminates the need for manual feature extraction and allows for the automatic learning of features. CNNs have proven effective in AD classification by leveraging the rich information provided by various imaging modalities. Several techniques have been proposed, including hybrid deep learning frameworks, improved CNN architectures, multi-modality cascaded networks, and multi-scale feature fusion networks. Deep learning methods, such as stacked deep polynomial networks and generative adversarial networks (GANs), have also been explored. Moreover, 2D and 3D CNNs have been employed, with some models incorporating pre-trained feature extractors and utilising techniques like Lasso, PCA, and landmark detection. Additionally, structural MRI methods, automatic segmentation methods, and multimodal fusion techniques have contributed to improving AD diagnosis. These advancements offer valuable insights and tools for the classification and detection of AD using deep learning and multimodal brain imaging data. Addressing these research gaps is crucial to advancing the field of AD classification and improving the accuracy and applicability of deep learning models in clinical settings. By addressing issues related to generalization, interpretability, data scarcity, standardization, and the integration of clinical information, researchers can

enhance the reliability and effectiveness of AD classification methods, ultimately leading to improved diagnosis and treatment of AD patients.

3. Methodology

The early diagnosis of Alzheimer's disease (AD) can be formulated as a mathematical problem involving the identification of reliable biomarkers and the development of accurate diagnostic models. Let X be the set of potential biomarkers, including neuroimaging data, genetic markers, and biomarker measurements. Let Y be the set of diagnostic outcomes, with $Y = \{0, 1\}$ representing healthy individuals and individuals with AD, respectively. Let D be the dataset consisting of N samples, where each sample $d_i = (x_i, y_i)$ represents the biomarker measurements and corresponding diagnostic outcome for the i_{th} individual. The goal is to learn a mapping function $f : X \rightarrow Y$ that accurately predicts the diagnostic outcome based on the biomarker measurements. Our proposed architecture includes two steps: the training of individual deep convolutional classifiers and ensembling them. As an end-to-end method, deep learning relies on a large amount of labelled data to obtain features from graphs. Medical data on Alzheimer's disease, in particular, are scarce. In order to meet the requirements of the deep network, we adopt the transfer learning method, which can not only reduce the amount of calculation in the model, but can also greatly improve model performance. The last layer of the fully connected network can be altered and omitted depending on the categorising of our classes, namely, cognitively normal and mild cognitive impairment. Hence, the final fully connected layer for this purpose would have only 2 neuron units instead of 1000 employed in the models for the ILSVRC challenge. The proposed research work model architecture combines features from 8 models.

Features from all models are retrieved in this case, and the final layers are then modified as necessary. Before adding the final 2 units, dropouts are also added. The final layer computes the classification probability of each output type using the Softmax function:

$$\sigma\left(\vec{z}\right)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}, \tag{1}$$

where σ corresponds to Softmax function, $\vec{z}$ is the input vector, e^{z_i} and e^{z_j} denote the standard exponential function applied to the input and output vectors, while K represents the number of classes within the multi-class classifier.

Consequently, the accuracy of individual models was improved. Secondly, 8 individual architectures after training individually on dementia images and saving the model predictions and ensembling using different weights for each model's predictions are employed. This proposed ensemble of classifiers ensures that each architecture can operate independently.

By integrating the ensembled learners, it often outperforms a single algorithm by determining the average to obtain a higher accuracy, as it is less sensitive to noise, outliers, or changes in sampling [45]. The majority of the labels predicted by each particular result of feature representation for a subject constitute the expected label. The projected label for a certain subject is the label that indicates the highest likelihood of these findings for the subject if there is a scenario in which a subject cannot be categorised when we combine two results of a single feature representation.

3.1. Dataset Pre-Processing

This experiment was performed on 113 subjects taken from OASIS-3: Longitudinal Multimodal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and AD archive [46]. The MRI images are available in NifTI format. A total of 247 and 410 MRI scans were used for cognitively normal and mild cognitive impairment patients respectively. T1-weighted structural sequences were processed on the axial plane of brain MRIs. The Clinical Dementia Rating (CDR) Scale was used to determine the dataset's dementia status; a score of 0 on the scale indicates cognitively normal, while a score of more than 0 but less

than 2 indicates mild impairment. Each image in the dataset is completely characterised, and complete details are given.

The first task with this dataset was converting the image files from NifTI to PNG. Performing this conversion resulted in 255 slices of the axial view of the brain for each subject, as shown in Figure 1. Multiple axial scans of the brain of each patient were extracted, particularly the images containing the hippocampus embedded deep into the temporal lobe and some areas which demonstrate a significant shrinkage in some parts of the brain, were fed as an input to the neural network. Multiple slices of MRIs were extracted from each subject for accurate prediction of the disease. The middle slices contain more tissue than the border slices. Since tissues are a biomarker for dementia, central slices aid the system in precisely classifying disorders that resemble AD. For the categorisation of AD, the slices along the border are less helpful and more deceptive since they contain more of the bones and skulls. The central slices are therefore utilised for training. This section goes into great detail on the dataset preparation.

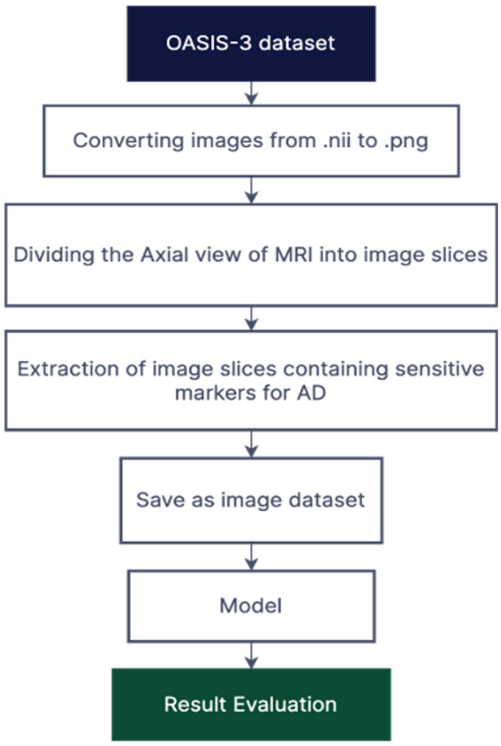


Figure 1. Illustration of the proposed methodology.

To make our predictions more robust, data augmentation technique was applied to our dataset. Since it is a classification model, the visual transformation mainly focused on flipping random images horizontally; rotating them by a certain degree and min-max normalisation; and rescaling inputs to the range of [0, 1], computed as follows:

$$y = \frac{(x - \min(x))}{(\max(x) - \min(x))}, \tag{2}$$

where $\min(x)$ and $\max(x)$ are the minimum and maximum intensity values, respectively, for the entire dataset, and y_i represents the normalised intensity value against location x_i (where $i \in [1, n]$). This helps our dataset to resolve the class imbalance issue, creating variations and generalizations for our model.

3.2. Ensemble Learning

Ensemble learning is a technique that combines the predictions of multiple individual models to make more accurate and robust predictions. It is widely used in machine

learning and has shown promising results in various domains, including Alzheimer's disease classification. In this paper, an ensemble of transfer learning models is proposed using two different methods: simple averaging and weighted averaging. By combining the predictions of these models, we aim to improve the extraction of sparse patterns and features from MRI images for Alzheimer's disease diagnosis. These learning models are used for identifying Alzheimer's disease. The proposed methodology's block diagram is shown in Figure 2. Our transfer learning models are input with 2D image slices.

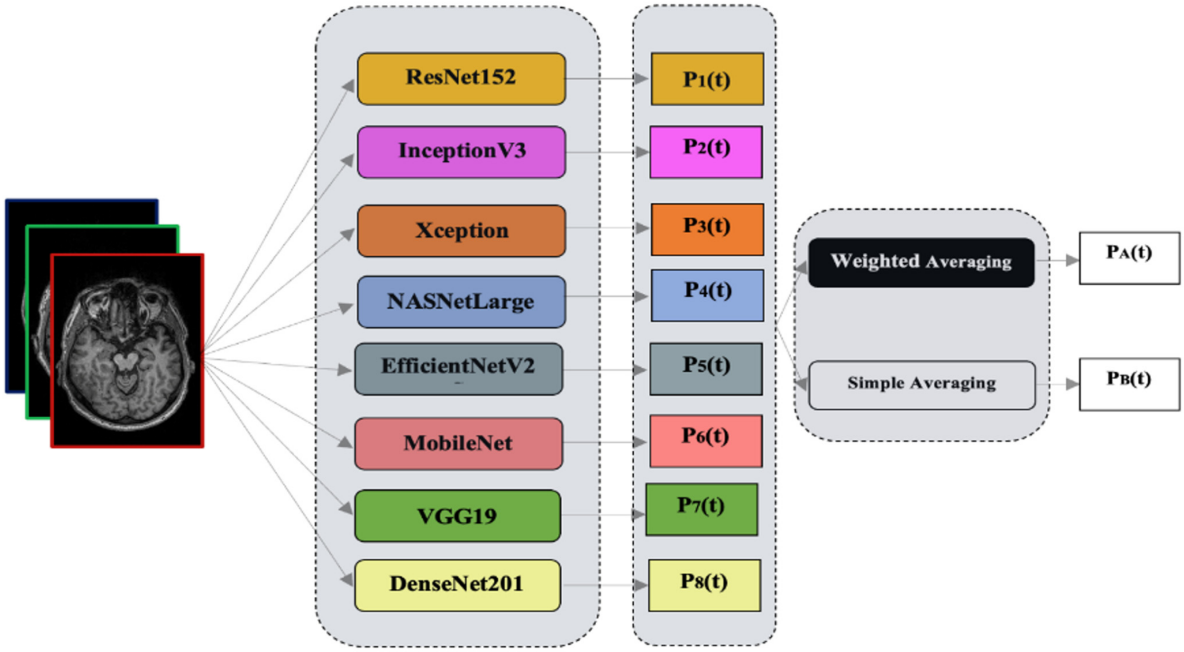


Figure 2. Graphical abstract of the proposed study.

Combining the judgements from several different models yields superior outcomes to those produced by a single model [47]. The generalization capacity of an ensemble, comprising multiple learners, is often considerably stronger than that of an individual learner, especially when it comes to weak learners. While an ensemble of weak learners can theoretically achieve good performance, in practice, people tend to prefer strong learners for various reasons. These reasons include minimizing the number of individual learners and leveraging the ability to reuse previously learned information about the strong learners [47].

In this paper, we utilise heterogeneous ensembles which have individual learners called component learners or simply individual learners. Unlike homogeneous ensembles, where all individual learners are of the same type, such as employing the same classification models, heterogeneous ensembles consist of diverse individual learners and learning algorithms, without relying on a single base learner or learning algorithm.

In order to compare and circumvent these problems, different ensemble models have been developed along with simple averaging. Weighted averaging provides various weights to various base learners. The goals were to determine which ensemble method had a strong correlation with the accuracy of the ensemble, which type of averaging had a high capacity to produce classification models that were not only accurate but also more diverse, and which types of ensembles were most likely to be able to enhance classification performance.

Simple Averaging: The most straightforward method to combine the output is by averaging them together. Let our ensemble contain T individual learners $\{h_1, h_2, h_3, \ldots\ldots, h_T\}$, where $h_i(x)$ is the result of h_i on x number of samples. The equation can be described as

$$H(x) = \frac{1}{T}\sum_{i=1}^{T} h_i(x), \tag{3}$$

This approach has the benefit being easily implemented [28], but it does not take into account that some models may perform better than others.

Weighted Averaging: The transfer learning model which has lower classification error is given a smaller weight when combined for ensembling, as described by [29]. The values of weights are assigned manually with the help of multiplying with a factor α to each model.

$$H(x) = \sum_{i=1}^{T} \alpha_i h_i(x), \tag{4}$$

where α is the weight of an individual learner h_i, and $\alpha_i \geq 0$, $\sum_{i=1}^{T} \alpha_i = 1$. It is shown in [30] that optimal choice for alpha is

$$\alpha_i = \frac{\sum_j C_{ij}{}^{-1}}{\sum_k \sum_j C_{kj}{}^{-1}}, \tag{5}$$

4. Experimentation and Results

4.1. Evaluation Indicators

The confusion matrix is a table that is commonly used in classification problems to evaluate the performance of a classification algorithm. It compares the actual and predicted values of a dataset and provides a summary of the number of true positives, true negatives, false positives, and false negatives for each class in a classification problem. The true positives (*TP*) are the number of correctly predicted positive instances, while the true negatives (*TN*) are the number of correctly predicted negative instances. The false positives (*FP*) are the number of negative instances that were incorrectly predicted as positive, and the false negatives (*FN*) are the number of positive instances that were incorrectly predicted as negative. Overall, the confusion matrix and its associated performance metrics provide valuable information for evaluating the performance of classification algorithms and identifying areas for improvement.

The following parameters and metrics are used for comparative analysis of data-driven models. In this study, the classifier correctly predicted *TP* and *TN* instances and incorrectly predicted *FP* and *FN* instances. The paper includes various performance metrics such as accuracy, sensitivity, specificity, and AUC score. It is a valuable tool for evaluating the performance of classification models and identifying areas where the model can be improved. To assess the unbiasing effectiveness of the proposed classification method, the *TP*, *FP*, *FN*, and *TN* statistical indices are computed and used to calculate accuracy, sensitivity, and specificity, as follows:

$$Accuracy = \frac{TP + FP}{TP + FP + TN + FN} \tag{6}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

The entire dataset contained 657 images. The models are trained and evaluated using 5-fold cross-validation technique. We used 30 epochs with a batch size of 16. The learning rate was fixed at 0.001 for all architectures with the Adam optimizer.

4.2. Model Selection

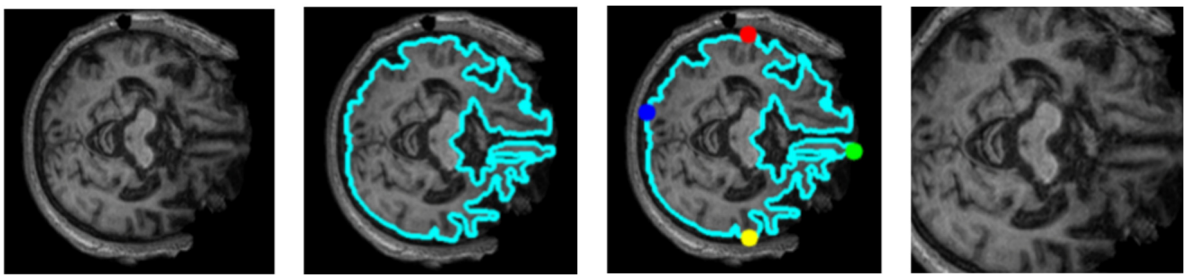
This section evaluates different pre-trained deep convolutional networks [48–55] on the OASIS dataset, as shown in Table 1. It can be observed that the VGG19 model performed the best with an accuracy of 0.977, whereas NASNetLarge model scored the least with

an accuracy of 0.912. MobileNet, DenseNet201, ResNet152, EfficientNetV2S, InceptionV3, and Xception scored an accuracy of 0.922, 0.961, 0.946, 0.952, 0.936, and 0.931, respectively. Furthermore, we improve the accuracy of the models using a cropping technique. The results demonstrate that DenseNet201 has a higher AUC score than VGG19, even though the accuracy of DenseNet201 is much less than that of VGG19. This discrepancy can be attributed to the fact that the AUC score only considers the ranking of the predicted probabilities. In contrast, the accuracy score considers both the ranking and the threshold for classification. In some cases, a model with a higher accuracy score may not have a better ROC curve and, thus, a lower AUC score.

Table 1. Performance of transfer learning models.

Models	Accuracy	Specificity	Sensitivity	AUC
MobileNet [48]	0.922	0.927	0.913	0.920
VGG19 [49]	0.977	0.949	0.961	0.958
DenseNet201 [50]	0.961	0.965	0.958	0.962
ResNet152 [51]	0.946	0.942	0.911	0.937
EfficientNetV2S [52]	0.952	0.966	0.943	0.959
InceptionV3 [53]	0.936	0.939	0.923	0.932
NASNetLarge [54]	0.912	0.927	0.913	0.920
Xception [55]	0.931	0.929	0.901	0.918

The cropping technique used extracted slices by finding the biggest contour, i.e., inside of the skull where the hippocampus is located. As shown in Figure 3, the extreme points of the brain were localised, which reduces the computation by exempting the pixels which do not contain the information of the features or unwanted space which may hinder the performance of classification model. The unwanted space also acts as noise for the image. To counter that, we applied threshold or canny edge detection, followed by finding the continuous points along the boundary having the same colour or intensity. The extreme points of the top, left, bottom, and right edges of the images were localised, and the same set of models with same hyperparameters as in without cropping the images were trained. As shown in Table 2, the performance of the eight models increased significantly in terms of accuracy and AUC score. It can be observed that VGG19 achieved the highest accuracy score of 0.982, and the lowest accuracy score was achieved by MobileNet. DenseNet201, ResNet152, EfficientNetV2S, InceptionV3, NASNetLarge, and Xception scored accuracies of 0.974, 0.966, 0.975, 0.952, 0.926, and 0.959, respectively. In terms of the AUC, VGG19 performed the best with a score of 0.986, whereas the lowest AUC score of 0.911 was achieved by NASNetLarge.



(**a**) (**b**) (**c**) (**d**)

Figure 3. Illustration of the proposed cropping technique for the extracting region of interest: (**a**) original image, (**b**) finding the biggest contour, (**c**) finding the extreme points, (**d**) region of interest.

Table 2. Performance of transfer learning models using cropping technique.

Models	Accuracy	Specificity	Sensitivity	AUC
MobileNet [48]	0.915	0.892	0.867	0.899
VGG19 [49]	0.982	0.973	0.993	0.986
DenseNet201 [50]	0.974	0.967	0.975	0.969
ResNet152 [51]	0.966	0.957	0.968	0.959
EfficientNetV2S [52]	0.975	0.988	0.969	0.973
InceptionV3 [53]	0.952	0.947	0.962	0.942
NASNetLarge [54]	0.926	0.899	0.901	0.911
Xception [55]	0.959	0.944	0.966	0.954

4.3. Model Ensembling Strategy

Training of these network was followed by ensembling in two ways i.e., simple averaging and weighted averaging. The experimentation settings for both the ensemble methods were kept same i.e., learning rate of 0.0001 with Adam optimizer and 100 Epochs. Images obtained through the cropping technique are used. Table 3 shows the performance of the two different ensemble methods. We select the best three models, i.e., VGG19, DenseNet201, and EfficientNetV2S according to Table 2 and denote their names as M1, M2, and M3 respectively. Next, we perform experiments by merging M1, M2, and M3 models in different combinations such as M1 + M2 + M3, M1 + M2, M2 + M3, and M1 + M3. According to Table 3, the best performance with an accuracy of 0.989 is achieved by combining models (M2 + M3) using simple averaging method. However, the lowest accuracy of 0.941 is achieved on M1 + M2 + M3 model that used weighted averaging method. In terms of the AUC, highest score is obtained by M1 + M2 with simple averaging method whereas lowest score is achieved by M1 + M2 + M3 model with weighted averaging method.

Table 3. Performance of ensemble networks using two different methods.

Ensemble Method	Models	Accuracy	Specificity	Sensitivity	AUC
Simple Average	M1 + M2 + M3	0.971	0.960	0.973	0.966
	M1 + M2	0.969	0.958	0.983	0.979
	M1 + M3	0.977	0.989	0.969	0.977
	M2 + M3	0.989	0.967	0.978	0.968
Weighted Average	M1 + M2 + M3	0.941	0.942	0.922	0.936
	M1 + M2	0.976	0.961	0.984	0.978
	M1 + M3	0.969	0.964	0.936	0.952
	M2 + M3	0.981	0.970	0.925	0.947

4.4. Visualizations

A feature map is produced as a result of the convolution operation being performed on the input by a convnet filter. To create an output volume, we apply a number of filters and stack the feature maps that are produced. Instead of looking at a single feature map, we have displayed multiple hidden feature maps in Figure 4 to show how the input is modified through subsequent layers. We exhibited eight VGG19 feature maps per layer for viewing purposes. As we move through different levels, there are several observations regarding the feature maps. Simple forms are picked up by the filters in the first layer. The hidden representation shows various cross-sectional anatomies of the brain, such as the hippocampus, cerebral artery, and different lobes. The majority of the information provided in the image is retained in the first layer feature maps, i.e., conv1 of block1. The first layers

in CNN architectures typically serve as edge detectors. The feature maps appear to be an abstract depiction of the original when we delve deeper into the network. Despite being less visible to humans, they still encode useful features. As we dig deeper, the feature maps become sparser because the filters pick up fewer features.

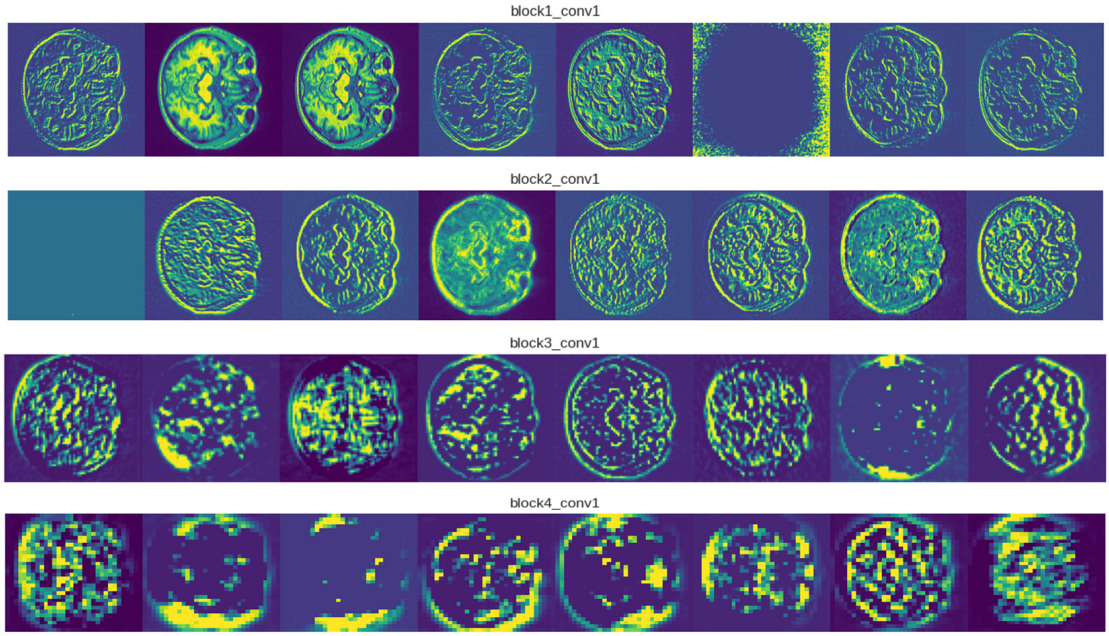


Figure 4. Visualizations for examples of MRI images from the OASIS-3 dataset.

5. Conclusions

In this paper, we proposed an ensemble-based transfer learning approach using two methods: simple averaging and weighted averaging. The proposed framework extracts superior sparse patterns and features from MRI images. To avoid overfitting, we used data augmentation techniques. After obtaining superior results from experiments, it can be inferred that both ensemble methods are equally competent. Furthermore, we investigated the performance of the models using uncropped and cropped MRI images. In future work, we aim to further refine and expand the proposed ensemble framework by incorporating additional transfer learning models, exploring different ensembling approaches, evaluating its performance on various AD databases, and leveraging adversarial networks for synthetic image generation. We also plan to test the results of our architecture on other AD databases to improve the generalization of our model. These advancements can potentially enhance the accuracy, robustness, and generalization of the model for Alzheimer's disease classification.

Author Contributions: Conceptualization, P.G., K.C. and M.P.; methodology, P.G., K.C. and X.Z.; software, P.G. and K.C.; validation, A.S., S.P., T.J. and M.P.; formal analysis, P.G. and X.Z.; investigation, P.G., K.C. and X.Z.; resources, A.S., S.P. and T.J.; data curation, P.G. and X.Z.; writing—original draft preparation, P.G.; writing—review and editing, P.G., K.C., X.Z., A.S., S.P., T.J. and M.P.; visualization, K.C., X.Z. and A.S.; supervision, M.P.; project administration, S.P. and T.J.; funding acquisition, T.J. and M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: We OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease [24].

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

References

1. Alzheimer's Association. 2022 Alzheimer's Disease Facts and Figures. *Alzheimer's Dement.* **2022**, *18*, 700–789. [CrossRef] [PubMed]
2. Thushara, A. An Efficient Alzheimer's Disease Prediction Based on MEPC-SSC Segmentation and Momentum Geo-Transient MLPs. *Comput. Biol. Med.* **2022**, *151*, 106247. [CrossRef] [PubMed]
3. Thapa, S.; Singh, P.; Jain, D.K.; Bharill, N.; Gupta, A.; Prasad, M. Data-Driven Approach Based on Feature Selection Technique for Early Diagnosis of Alzheimer's Disease. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
4. Adhikari, S.; Thapa, S.; Naseem, U.; Singh, P.; Huo, H.; Bharathy, G.; Prasad, M. Exploiting Linguistic Information from Nepali Transcripts for Early Detection of Alzheimer's Disease Using Natural Language Processing and Machine Learning Techniques. *Int. J. Hum. Comput. Stud.* **2022**, *160*, 102761. [CrossRef]
5. Tanveer, M.; Richhariya, B.; Khan, R.U.; Rashid, A.H.; Khanna, P.; Prasad, M.; Lin, C.T. Machine Learning Techniques for the Diagnosis of Alzheimer's Disease. *ACM Trans. Multimed. Comput. Commun. Appl.* **2020**, *16*, 1–35. [CrossRef]
6. Tremblay-Mercier, J.; Madjar, C.; Das, S.; Pichet Binette, A.; Dyke, S.O.M.; Étienne, P.; Lafaille-Magnan, M.-E.; Remz, J.; Bellec, P.; Louis Collins, D.; et al. Open Science Datasets from PREVENT-AD, a Longitudinal Cohort of Pre-Symptomatic Alzheimer's Disease. *NeuroImage Clin.* **2021**, *31*, 102733. [CrossRef] [PubMed]
7. Poulin, S.P.; Dautoff, R.; Morris, J.C.; Barrett, L.F.; Dickerson, B.C. Amygdala Atrophy Is Prominent in Early Alzheimer's Disease and Relates to Symptom Severity. *Psychiatry Res. Neuroimaging* **2011**, *194*, 7–13. [CrossRef] [PubMed]
8. Westman, E.; Cavallin, L.; Muehlboeck, J.-S.; Zhang, Y.; Mecocci, P.; Vellas, B.; Tsolaki, M.; Kłoszewska, I.; Soininen, H.; Spenger, C.; et al. Sensitivity and Specificity of Medial Temporal Lobe Visual Ratings and Multivariate Regional MRI Classification in Alzheimer's Disease. *PLoS ONE* **2011**, *6*, e22506. [CrossRef]
9. Gupta, A.; Kumar, D.; Verma, H.; Tanveer, M.; Javier, A.P.; Lin, C.-T.; Prasad, M. Recognition of Multi-Cognitive Tasks from EEG Signals Using EMD Methods. *Neural Comput. Appl.* **2022**. [CrossRef]
10. Kiani, M.; Andreu-Perez, J.; Hagras, H.; Papageorgiou, E.I.; Prasad, M.; Lin, C.-T. Effective Brain Connectivity for FNIRS With Fuzzy Cognitive Maps in Neuroergonomics. *IEEE Trans. Cogn. Dev. Syst.* **2022**, *14*, 50–63. [CrossRef]
11. Ding, W.; Lin, C.-T.; Prasad, M. Hierarchical Co-Evolutionary Clustering Tree-Based Rough Feature Game Equilibrium Selection and Its Application in Neonatal Cerebral Cortex MRI. *Expert Syst. Appl.* **2018**, *101*, 243–257. [CrossRef]
12. Lazli, L.; Boukadoum, M.; Mohamed, O.A. A Survey on Computer-Aided Diagnosis of Brain Disorders through MRI Based on Machine Learning and Data Mining Methodologies with an Emphasis on Alzheimer Disease Diagnosis and the Contribution of the Multimodal Fusion. *Appl. Sci.* **2020**, *10*, 1894. [CrossRef]
13. Chen, Z.S.; Kulkarni, P.; Galatzer-Levy, I.R.; Bigio, B.; Nasca, C.; Zhang, Y. Modern Views of Machine Learning for Precision Psychiatry. *Patterns* **2022**, *3*, 100602. [CrossRef] [PubMed]
14. Zhang, Y.; Zhang, H.; Adeli, E.; Chen, X.; Liu, M.; Shen, D. Multiview Feature Learning With Multiatlas-Based Functional Connectivity Networks for MCI Diagnosis. *IEEE Trans. Cybern.* **2022**, *52*, 6822–6833. [CrossRef]
15. Anh, N.; Prasad, M.; Srikanth, N.; Sundaram, S. Wave Forecasting Using Meta-Cognitive Interval Type-2 Fuzzy Inference System. *Procedia Comput. Sci.* **2018**, *144*, 33–41. [CrossRef]
16. Za'in, C.; Pratama, M.; Prasad, M.; Puthal, D.; Lim, C.P.; Seera, M. Motor Fault Detection and Diagnosis Based on a Meta-Cognitive Random Vector Functional Link Network. In *Fault Diagnosis of Hybrid Dynamic and Complex Systems*; Springer International Publishing: Cham, Switzerland, 2018; pp. 15–44.
17. Castellazzi, G.; Cuzzoni, M.G.; Cotta Ramusino, M.; Martinelli, D.; Denaro, F.; Ricciardi, A.; Vitali, P.; Anzalone, N.; Bernini, S.; Palesi, F.; et al. A Machine Learning Approach for the Differential Diagnosis of Alzheimer and Vascular Dementia Fed by MRI Selected Features. *Front. Neuroinform.* **2020**, *14*. [CrossRef] [PubMed]
18. Battineni, G.; Chintalapudi, N.; Amenta, F.; Traini, E. A Comprehensive Machine-Learning Model Applied to Magnetic Resonance Imaging (MRI) to Predict Alzheimer's Disease (AD) in Older Subjects. *J. Clin. Med.* **2020**, *9*, 2146. [CrossRef] [PubMed]
19. Alickovic, E.; Subasi, A. *Automatic Detection of Alzheimer Disease Based on Histogram and Random Forest*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 91–96.
20. Bandyopadhyay, A.; Ghosh, S.; Bose, M.; Singh, A.; Othmani, A.; Santosh, K. *Alzheimer's Disease Detection Using Ensemble Learning and Artificial Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 12–21.
21. Wang, H.; Shen, Y.; Wang, S.; Xiao, T.; Deng, L.; Wang, X.; Zhao, X. Ensemble of 3D Densely Connected Convolutional Network for Diagnosis of Mild Cognitive Impairment and Alzheimer's Disease. *Neurocomputing* **2019**, *333*, 145–156. [CrossRef]
22. Ortiz, A.; Munilla, J.; Górriz, J.M.; Ramírez, J. Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease. *Int. J. Neural Syst.* **2016**, *26*, 1650025. [CrossRef]
23. Tanveer, M.; Rashid, A.H.; Ganaie, M.A.; Reza, M.; Razzak, I.; Hua, K.-L. Classification of Alzheimer's Disease Using Ensemble of Deep Neural Networks Trained Through Transfer Learning. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 1453–1463. [CrossRef]
24. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *arXiv* **2013**, arXiv:1311.2524.
25. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How Transferable Are Features in Deep Neural Networks? *arXiv* **2014**, arXiv:1411.1792.
26. Balaji, P.; Chaurasia, M.A.; Bilfaqih, S.M.; Muniasamy, A.; Alsid, L.E.G. Hybridized Deep Learning Approach for Detecting Alzheimer's Disease. *Biomedicines* **2023**, *11*, 149. [CrossRef] [PubMed]

27. Pei, Z.; Wan, Z.; Zhang, Y.; Wang, M.; Leng, C.; Yang, Y.-H. Multi-Scale Attention-Based Pseudo-3D Convolution Neural Network for Alzheimer's Disease Diagnosis Using Structural MRI. *Pattern Recognit.* **2022**, *131*, 108825. [CrossRef]
28. Liu, M.; Cheng, D.; Wang, K.; Wang, Y. Multi-Modality Cascaded Convolutional Neural Networks for Alzheimer's Disease Diagnosis. *Neuroinformatics* **2018**, *16*, 295–308. [CrossRef] [PubMed]
29. Abdou, M.A. Literature Review: Efficient Deep Neural Networks Techniques for Medical Image Analysis. *Neural Comput. Appl.* **2022**, *34*, 5791–5812. [CrossRef]
30. Zhang, T.; Shi, M. Multi-Modal Neuroimaging Feature Fusion for Diagnosis of Alzheimer's Disease. *J. Neurosci. Methods* **2020**, *341*, 108795. [CrossRef]
31. Odusami, M.; Maskeliūnas, R.; Damaševičius, R.; Misra, S. Explainable Deep-Learning-Based Diagnosis of Alzheimer's Disease Using Multimodal Input Fusion of PET and MRI Images. *J. Med. Biol. Eng.* **2023**, *43*, 291–302. [CrossRef]
32. Yu, W.; Lei, B.; Ng, M.K.; Cheung, A.C.; Shen, Y.; Wang, S. Tensorizing GAN With High-Order Pooling for Alzheimer's Disease Assessment. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 4945–4959. [CrossRef]
33. Shi, J.; Zheng, X.; Li, Y.; Zhang, Q.; Ying, S. Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 173–183. [CrossRef]
34. Zhang, X.; Yao, L.; Wang, X.; Monaghan, J.; McAlpine, D.; Zhang, Y. A Survey on Deep Learning-Based Non-Invasive Brain Signals: Recent Advances and New Frontiers. *J. Neural Eng.* **2021**, *18*, 031002. [CrossRef]
35. Suk, H.-I.; Lee, S.-W.; Shen, D. Deep Ensemble Learning of Sparse Regression Models for Brain Disease Diagnosis. *Med. Image Anal.* **2017**, *37*, 101–113. [CrossRef] [PubMed]
36. Feng, W.; Van Halm-Lutterodt, N.; Tang, H.; Mecum, A.; Mesregah, M.K.; Ma, Y.; Li, H.; Zhang, F.; Wu, Z.; Yao, E.; et al. Automated MRI-Based Deep Learning Model for Detection of Alzheimer's Disease Process. *Int. J. Neural Syst.* **2020**, *30*, 2050032. [CrossRef] [PubMed]
37. Wang, S.; Wang, H.; Shen, Y.; Wang, X. Automatic Recognition of Mild Cognitive Impairment and Alzheimers Disease Using Ensemble Based 3D Densely Connected Convolutional Networks. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 517–523.
38. Jain, R.; Jain, N.; Aggarwal, A.; Hemanth, D.J. Convolutional Neural Network Based Alzheimer's Disease Classification from Magnetic Resonance Brain Images. *Cogn. Syst. Res.* **2019**, *57*, 147–159. [CrossRef]
39. Zhu, W.; Sun, L.; Huang, J.; Han, L.; Zhang, D. Dual Attention Multi-Instance Deep Learning for Alzheimer's Disease Diagnosis With Structural MRI. *IEEE Trans. Med. Imaging* **2021**, *40*, 2354–2366. [CrossRef]
40. Chen, Y.; Shi, B.; Wang, Z.; Zhang, P.; Smith, C.D.; Liu, J. Hippocampus Segmentation through Multi-View Ensemble ConvNets. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 192–196.
41. Ataloglou, D.; Dimou, A.; Zarpalas, D.; Daras, P. Fast and Precise Hippocampus Segmentation Through Deep Convolutional Neural Network Ensembles and Transfer Learning. *Neuroinformatics* **2019**, *17*, 563–582. [CrossRef]
42. Zhang, D.; Shen, D. Semi-Supervised Multimodal Classification of Alzheimer's Disease. In Proceedings of the 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, 30 March–2 April 2011; pp. 1628–1631.
43. Liu, S.; Liu, S.; Cai, W.; Che, H.; Pujol, S.; Kikinis, R.; Feng, D.; Fulham, M.J. ADNI Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1132–1140. [CrossRef]
44. Liu, C.; Huang, F.; Qiu, A. Monte Carlo Ensemble Neural Network for the Diagnosis of Alzheimer's Disease. *Neural Netw.* **2023**, *159*, 14–24. [CrossRef] [PubMed]
45. Yang, Y.; Li, X.; Wang, P.; Xia, Y.; Ye, Q. Multi-Source Transfer Learning via Ensemble Approach for Initial Diagnosis of Alzheimer's Disease. *IEEE J. Transl. Eng. Health Med.* **2020**, *8*, 1–10. [CrossRef]
46. LaMontagne, P.J.; Benzinger, T.L.S.; Morris, J.C.; Keefe, S.; Hornbeck, R.; Xiong, C.; Grant, E.; Hassenstab, J.; Moulder, K.; Vlassenko, A.G.; et al. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease. *medRxiv* **2019**. [CrossRef]
47. Ganaie, M.A.; Hu, M.; Malik, A.K.; Tanveer, M.; Suganthan, P.N. Ensemble Deep Learning: A Review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105151. [CrossRef]
48. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
49. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
50. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* **2016**, arXiv:1608.06993.
51. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
52. Tan, M.; Le, Q.V. EfficientNetV2: Smaller Models and Faster Training. *arXiv* **2021**, arXiv:2104.00298.
53. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv* **2015**, arXiv:1512.00567.

54. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V. Learning Transferable Architectures for Scalable Image Recognition. *arXiv* **2017**, arXiv:1707.07012.
55. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv* **2016**, arXiv:1610.02357.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Clean Technologies and Environmental Policy
https://doi.org/10.1007/s10098-023-02596-y

REVIEW

Environmental sustainability analysis of biofuels: a critical review of LCA studies

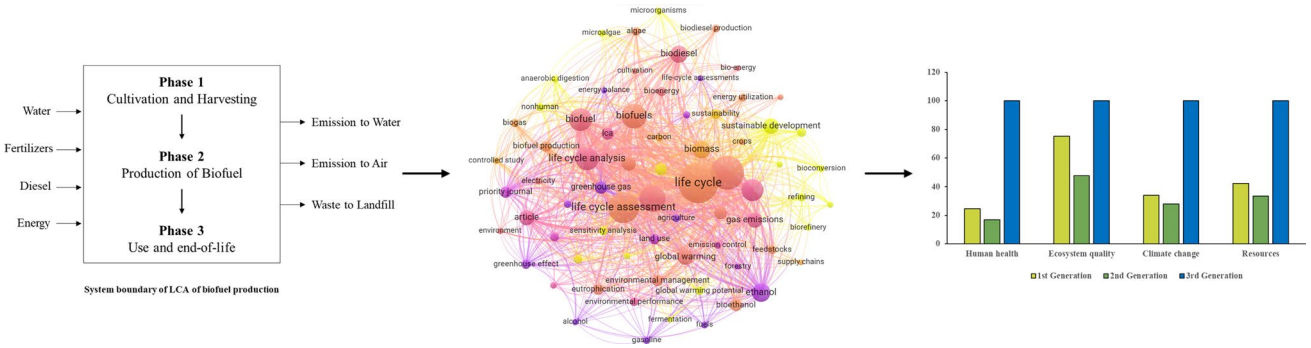
Kulvendra Patel[1] · **S. K. Singh**[1]

Received: 26 April 2023 / Accepted: 10 August 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The use of fossil fuels in the transportation sector has significantly increased the greenhouse gas (GHG) emissions, leading to concerns about their environmental impact. As a result, biofuels have been promoted as a promising alternative to fossil fuels to mitigate GHG emissions. However, while biofuels are considered a low-carbon solution, there are concerns that its widespread use could have unforeseen environmental effects. This study analyzes the intricacies involved in biofuel's life cycle assessment (LCA). Fifty-three peer-reviewed studies were examined, covering a broad range of biofuels produced from over 25 different feedstocks. The literature review focused mainly on research published from 2012 to 2022 to avoid outdated information. The study found that second-generation biofuels have a greater potential for reducing GHG emissions than first-generation biofuels. However, the GHG emissions from third-generation biofuels are higher than those from conventional fuel, ranging from 10.2 to 1910 g CO_2 eq. / MJ. The net energy ratio for biofuels ranges from 1.23 to 12.49 for first-generation, 0.003 to 15.04 for second-generation, and 3 to 18.5 for third-generation biofuels. The analysis also shows that the water footprint of biofuels is 50–240 times greater than fossil fuels. The study also examines the critical methodological aspects and sources of uncertainty in the LCA of biofuels and suggests solutions to address these problems. Overall, the study highlights the importance of taking a comprehensive approach when assessing the environmental impact of biofuels and the need for continued research to develop sustainable and effective biofuels.

Graphical abstract



Keywords Biofuels · Environmental impact assessment · Global warming potential · Life cycle assessment · Sustainability

Introduction

Energy production and use have been related to urbanization, modernization, and industrialization in various economic, scientific, and social fields. The depletion of non-renewable fuels and the subsequent rise in GHG emissions has become

✉ Kulvendra Patel
kulvendrapatel_2k21phden501@dtu.ac.in

[1] Department of Environmental Engineering, Delhi Technological University, Delhi 110042, India

Published online: 23 August 2023

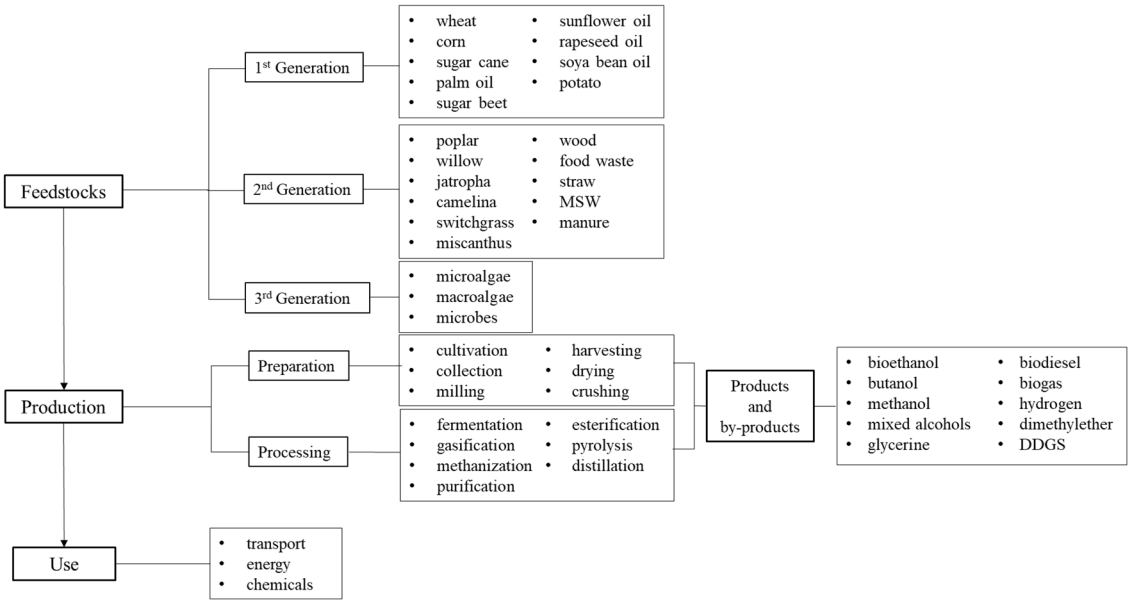


Fig. 1 Feedstock and production processes for different biofuels

a critical problem (Fawzy et al. 2020; Osman et al. 2021a). Economic development is accelerating, particularly in developing nations like India and China. These nations have now reached the stage of economic development when energy use is at its highest due to industrialization. Meeting the rising energy demand while having sustainable energy supplies for the future is thus a significant problem (Yazan et al. 2017). Between 1971 and 2019, the total energy supply worldwide increased 2.6 times, from 230 to 606 EJ. The primary energy sources for power generation in 2019 were still oil and coal, which accounted for 31% and 27% of global energy production, respectively (IEA 2021). The Intergovernmental Panel on Climate Change (IPCC) Working Group III (WG3) has identified energy systems, industry, buildings, transportation, and AFOLU (agriculture, forestry, and other land uses) as the five primary sectors responsible for global greenhouse gas emissions (Lamb et al. 2021). GHG emissions from transportation have been growing faster compared to other industries. The leading cause of GHG emissions related to transportation is the burning of fossil fuels in various modes of transport. The fuel used for transportation mainly consists of petroleum, which makes up over 90% of gasoline and diesel. Globally, transportation accounts for 15% of total GHG emissions and 23% of CO_2 emissions associated with energy consumption (Sims and Schaeffer 2014; Uusitalo et al. 2014; Jeswani et al. 2020). Biofuels are commonly recognized as possible substitute fuels for transportation that can decrease reliance on fuels made from petroleum and help address the problem of climate change. As an alternative to conventional fuels, several feedstocks and conversion methods for biofuel have been proposed.

Biofuels refer to energy-enriched compounds derived from biomass, such as biodegradable waste or residue from forestry, agriculture, fishing, or biodegradable urban or industrial waste. The kind of feedstock, the conversion method, the technical specifications of the fuel, and its intended application are a few distinguishing factors that may be used to categorize biofuels. Different definitions for different biofuel types are in use due to the vast number of potential differences (Bharti et al. 2021). Biofuel sources are categorized into three generations of feedstock: edible energy and oil-based crops (first generation), lignocellulose-based (second generation) (Chen et al. 2019), and algae-based (third generation)(Kumar Sharma et al. 2021; Ocreto et al. 2021). Figure 1 provides an outline of several biofuel types, feedstocks, and conversion pathways.

Systems for producing, converting, and supplying sustainable renewable energy must be effective and sustainable in terms of the environment, the economy, and society for both the present and future. Identifying biofuel sources can potentially lessen the environmental effect of fossil fuels (Balajii and Niju 2019; Gunarathne et al. 2019). Biofuels can also help mitigate the rising fossil fuel usage and relieve dependence on non-renewable resources (Hassan et al. 2020; Peng et al. 2020). How to define sustainable biomass production is a topic of debate on the national and international levels (Kaloudas et al. 2021). First and foremost, biofuel should lower energy use and improve energy security by using fewer petroleum-based products. Second, it must be advantageous for the environment by lowering greenhouse gas emissions, conserving diverse land uses, protecting soil productivity, water quality and quantity, and biodiversity.

Third, it should not unnecessarily affect food availability and other resources while improving ecosystem functions. It should also be competitive economically and contribute more energy than it uses in production (BRDI 2008; Munagala et al. 2022). Bioethanol and biodiesel are popular biofuels for blending with Petro-diesel to lessen reliance on traditional diesel. Bioethanol production involves the conversion of starch or biomass with high sugar content (like sugar cane or grains) into sugar, which is then fermented and distilled. Another method for producing bioethanol or bio-oil is through hydrolysis of lignocellulosic biomass, followed by fermentation and distillation, also referred to as biomass to liquid (BTL). Alternatively, biodiesel can be produced by using alcohol to extract and esterify vegetable oils, discarded cooking oils, and animal fats (Sinha et al. 2013; Foteinis et al. 2020). India is attempting to embrace renewable energy sources, including biofuel, to reduce its reliance on conventional energy. Since utilizing green fuels like biodiesel and ethanol instead of traditional transportation fuels helps reduce greenhouse gas emissions, these renewable fuels are being advocated globally. Some of the research that suggests that biofuel production using a biorefinery platform might lower carbon emissions is shown in Table 1. Brazil is the global leader with a fleet of vehicles that uses gasoline-ethanol blends with a 27% content (USDA 2021a). In all of its fleet vehicles, the United States now uses roughly 7.1% biofuels, while the European Union (EU) has a bold ambition to utilize 10% renewable fuels by 2020 (EASAC 2012). India has stepped up its efforts to achieve its objective of being in the E-20 by Ethanol Supply Year (ESY) 2025. While maintaining its immediate objective of E-10 by ESY 2022, the Indian government stated its E-20 target year of ESY 2025 instead of 2030 (USDA 2021b).

The sustainability and benefits of biofuels are impacted by the emissions and energy used in their manufacturing process. Therefore, to accurately assess the advantages of using biofuels instead of traditional fossil fuels, it is essential to use reliable, practical, and scientific evaluation methods (Chamkalani et al. 2020; Kargbo et al. 2021). To evaluate the reductions in emissions by utilizing biofuels in India, a life cycle assessment (LCA) must be performed. The LCA evaluates the complete life cycle of a product or process, including the acquisition of raw material through manufacturing, usage, and disposal, to measure its environmental impact by creating a comprehensive inventory of all the material inputs, energy consumption, and emissions associated with the product or process. The ultimate goal of LCA is to provide policymakers with a better understanding of the environmental impact by analyzing the assessment results (Soam et al. 2015). The life cycle assessment method will light India's energy and environmental advantages of biofuel production. Understanding the GHG emission reduction that biofuels may achieve compared to their fossil fuel counterparts is crucial for guiding policy development and selecting optimal fuel choices (RAE 2017). By calculating their carbon footprint, several studies have examined the biofuel's potential to reduce GHG emissions. However, their conclusions are sometimes contradictory because of the vast range of estimations. In order to influence future policy, the primary goal of this study is to increase clarity and knowledge of the carbon footprints of various liquid biofuels (Khanna et al. 2011). The study evaluates the techniques and environmental sustainability of biofuels by examining the prior LCA studies, highlighting significant methodological approaches, and presenting critical conclusions.

Bibliometric analysis

To locate pertinent publications and pinpoint areas where research on the environmental effects of biofuel production is lacking, a search strategy using Boolean operators was employed in the WoS core collection database. The bibliometric mapping data shown in Figure S1a (Online Resource) were obtained from the WoS and visualized using VOSviewer software to show the co-occurrence of keywords in the literature from 2012 to 2022. Bibliometric research methodology: TOPIC: "life cycle assessment" AND TOPIC: 'Biofuels' OR TOPIC: 'Bioethanol' OR TOPIC: 'Biodiesel'

Table 1 Carbon emission reduction from various feedstocks during the biofuel production

Feedstock	Products & Co-products	Reduction in CO_2 emissions	References
Sugarcane	Bioethanol, Biogas, Electricity, steam, and power	40–70%	(Gnansounou et al. 2015; Silalertruksa et al. 2017)
Palm oil	Biodiesel	20–25%	(Lim and Lee 2011; Lee and Ofori-Boateng 2013)
Corn stover & wheat straw	Bioenergy	47–50%	(Cherubini and Ulgiati 2010; Gallejones et al. 2015)
Forest residue	Biodiesel, Biogas, Electricity	51–80%	(Karlsson et al. 2014; Spatari et al. 2020)
Jatropha, Sorghum	Biodiesel	75–80%	(Kumar et al. 2012; Vrech et al. 2019; Martinez-Hernandez et al. 2019)
Rapeseed	Biodiesel, Energy	30–50%	(Fernández-Tirado et al. 2016; Gupta et al. 2022)
Used cooking oil	Biodiesel	30–55%	(Vrech et al. 2019; Bhonsle et al. 2022)

OR TOPIC: 'Lignocellulosic.' The bibliographic search was limited to articles and reviews, and only documents published between 2012 and 2022 were taken into account. Over ten years (2012–2022), bibliometric mapping shows that terms such as life cycle assessment, environmental impacts, biofuel, carbon dioxide, and greenhouse effect dominate the literature. The density visualization mapping shown in Figure S1b (Online Resource) shows that the literature focused on environmental impacts. The darker region indicates the focused area, and the lighter region shows the research gap in the study area that needs further research in the near future. For example, greater research is needed on social life cycle assessment, water footprint, waste management, bioeconomy, land use change, and techno-economic analysis.

The bibliometric analysis shows a diverse geographical distribution of research. The United States takes the lead, accounting for 24% of the studies, followed by China at 12%. Europe also emerges as a significant contributor, with noteworthy contributions from countries like the UK (8%), Italy (6%), Spain (6%), Germany (5%), and France (3%). Moreover, emerging economies such as Brazil (8%) and India (7%) demonstrate a growing interest in this field, showcasing significant research efforts, as shown in Figure S2 and Figure S3 (Online Resource). These findings highlight the global reach and interest in LCA studies of biofuels across different regions. The network map of keyword co-occurrence among 53 research is shown in Figure S4 (Online Resource) and was created using the VOSViewer (Perianes-Rodriguez et al. 2016) full counting approach.

Methodology and framework

Literature search and selection criteria

We performed a literature search in different databases, including Web of Science (WoS), Scopus, Science Direct, and PubMed, to identify peer-reviewed studies on the sustainability of biofuels. The publications in this study were selected if they (1) revealed that LCA was performed, (2) had well-defined system boundaries, and (3) examined a set of impact (midpoint or damage) categories. A qualitative analysis was conducted once all pertinent research had been considered. The literature search mainly concentrated on studies published between 2012 and 2022 to minimize obsolete material. A total of 53 articles were selected for analysis in the present study, including the life cycle analysis of biofuels, the geographic location, functional unit, feedstock used, impact assessment methodology, system boundaries, and findings. The articles consist of different kinds of biofuels from various feedstocks, classified into first, second, and third generations, with over 25 types of feedstocks being used. Table 2 shows an overview of the evaluated studies on feedstock, life cycle tools, impact methods, and functional units. Among the selected articles, 26% evaluated first-generation biofuels, 57% evaluated second-generation biofuels, and 17% evaluated third-generation biofuels.

Life cycle assessment

LCA is a method or procedure used to evaluate the potential environmental effects of a product, process, or activity by considering every step of the products and the raw materials supply chains, from production to final disposal. It was carried out in accordance with the ISO 14040:2006 and 14044:2006 standard framework to ensure that exact conclusions were achieved (Wang et al. 2013a; Chau et al. 2015; Patel and Singh 2022a). An LCA study might be from "cradle to grave" or "cradle to gate," depending on the goal and scope (Khoo et al. 2019). Figure 2a shows the generalized system boundaries for the production of biofuels. Phase 1 includes cultivation, use of water, diesel, fertilizers, and air and water emissions due to the cultivation and harvesting process. Phase 2 includes the conversion of biomass to biofuels through various processes, the use of energy, and the associated emissions to air and water. Phase 3 incorporates biofuel and co-product transportation and emissions due use of biofuel. According to ISO 14,040, LCA consists of four phases: (i) goal and scope; (ii) life cycle inventory (LCI); (iii) life cycle impact assessment (LCIA); and (iv) interpretation.

Goal and scope definition

This initial step in every LCA study critical for understanding general outcomes and LCA findings, including defining the functional units and system boundaries (ISO 2006a; Harris et al. 2016). Figure 2b shows the percentage of reviewed studies considering different phases of the system boundary. All the studies have considered Phase 2 of the LCA system boundary, and only 14 studies have considered all three phases in the reviewed studies. This suggests that only a small number of studies have considered the use phase of biofuel and compared them with conventional fuels, indicating a significant gap in the LCA application for decision-making. The functional unit (FU) acts as a reference point for linking inputs and outputs. This reference point is essential to ensure that LCA outcomes are comparable; this is especially critical when comparing different systems and ensuring that comparisons are made consistently (ISO 2006bMartínez-Blanco et al. 2011; Patel and Singh 2022b). Figure 3 shows the percentage of reviewed studies considering different FUs with land use change (LUC) and without LUC analysis. It was observed that 22 studies had used biomass production as FU (kg or ton of biofuel produced), 18 studies had considered the unit of bioenergy (MJ, GJ, or

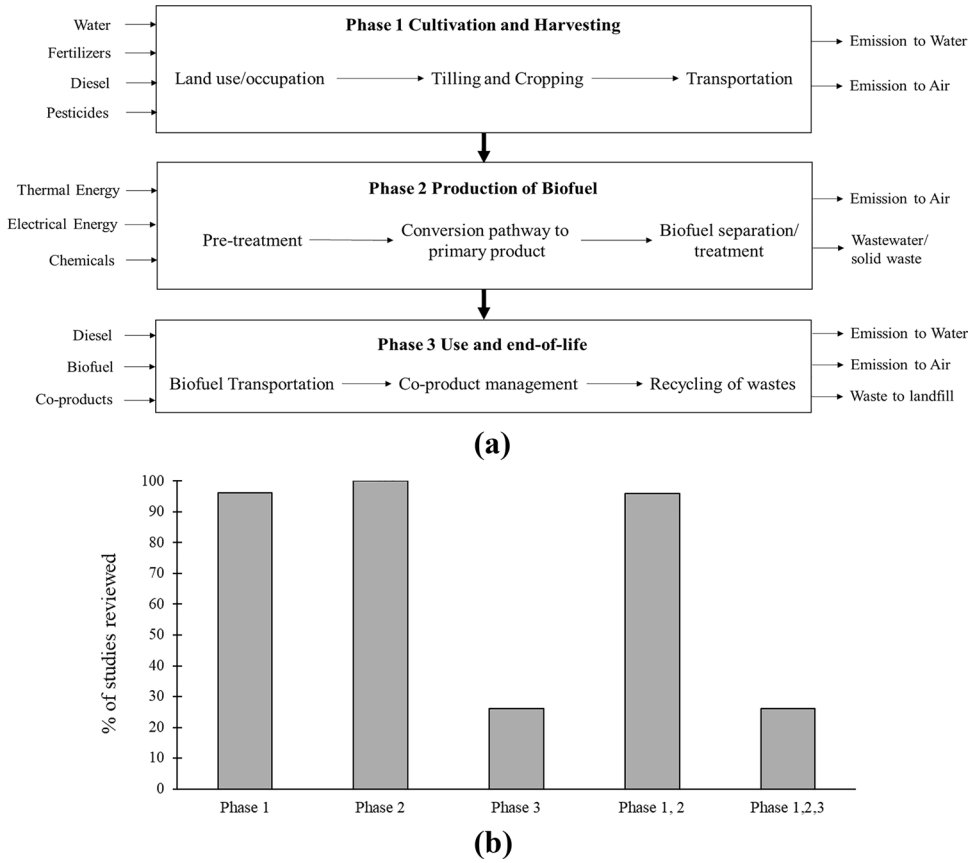


Fig. 2 **a** General system boundaries of LCA of biofuel production, **b** Phases of LCA for production of biofuels considered in the reviewed studies

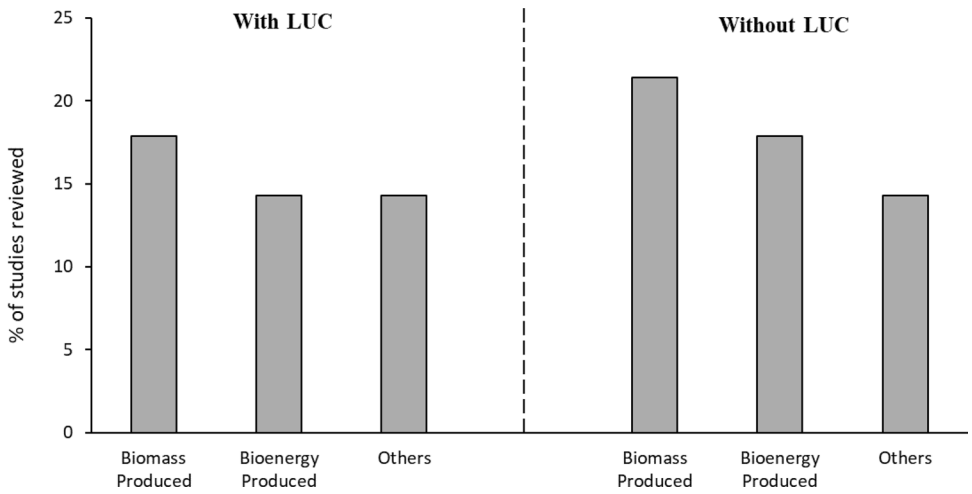


Fig. 3 Types of functional units used in reviewed LCA studies

kWh of energy produced) production as FU and 16 studies had used other FUs (Litre of biofuel, per km of car driven, and per hectare-area) (Borrion et al. 2012; Sreekumar et al. 2020; Isler-Kaya and Karaosmanoglu 2022).

Life cycle inventory and life cycle impact assessment

The LCI methodology includes gathering and evaluating information regarding the resources used, such as materials, energy, and water, and the emissions generated at every phase of the production of biofuels. LCI provides a detailed inventory of inputs and outputs of the biofuel production process (Chung et al. 2019; Chopra et al. 2020), while LCIA evaluates the potential environmental impacts associated with these inputs and outputs, including emissions from the cultivation of feedstocks, transportation, and processing of biofuel and helps in understanding the results of LCI (Rocha et al. 2014).

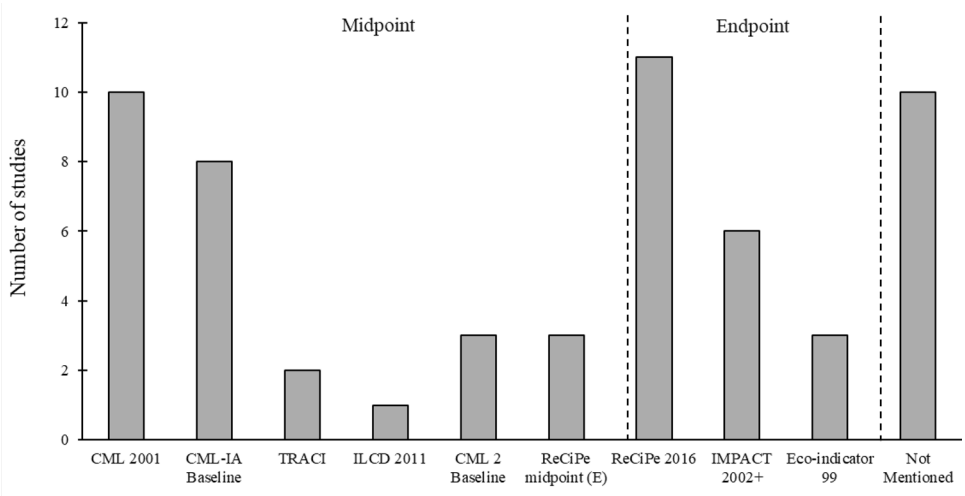


Fig. 4 Frequencies of LCIA methods in reviewed studies

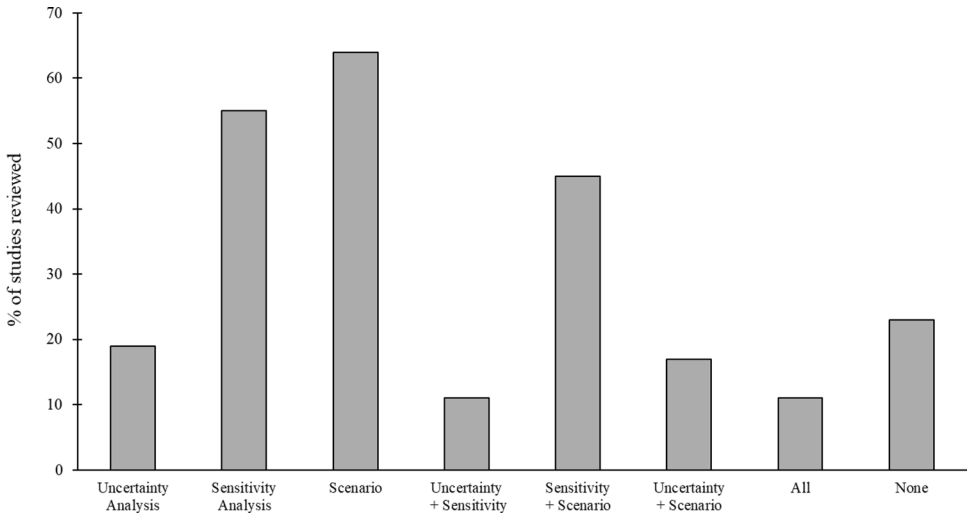


Fig. 5 Details of uncertainty, scenario, and sensitivity analysis in the reviewed studies

The database used for inventory analysis is shown in Table 2. IMPACT 2002+, ReCiPe 2016, EDIP 2003, CML, Ecopoint 97, Eco-indicator 99, and TRACI are the standard methods used to evaluate environmental impacts (ILCD 2010; Owsianiak et al. 2014). Figure 4 shows the no. of different impact methods used in reviewed studies and grouped into midpoint and endpoint impact categories.

Uncertainties and sensitivity analysis

LCA studies are models that are simplified representations of real-world systems and are thus fundamentally uncertain (Wang and Shen 2013). LCA includes trade-offs between numerous competing factors and, sometimes, uncertainty in relevant aspects (Guinée et al. 2017). The distribution of different environmental emissions, resource consumption, or effect categories may represent one of these uncertainties. The LCIA can also lead to uncertainty in the result. Endpoint impact categories have significantly more uncertainty but can produce more intelligible findings (Bare et al. 2000), whereas the midpoint method is less scientifically challenging and has much lower uncertainty (Gnansounou and Kenthorai Raman 2016). As shown in Fig. 5, only 19% of studies have done uncertainty analysis. These uncertainties can be reduced by using more precise data (Michiels and Geeraerd 2020; Ubando et al. 2022). The uncertainty for each effect category is calculated by dividing the standard deviation by the mean (Sharma et al. 2020). The calculation of uncertainties is performed on a study-specific basis. Each included study has its own mean values and corresponding standard deviations determined by the data and methodology employed in that particular study. Consequently, the uncertainties vary depending on the specific study and the effect category being considered. It considers uncertainties in data, models, and assumptions used in the analysis. The goal is to provide a measure of confidence or reliability in the reported results. Uncertainty analysis involves identifying sources of uncertainty, estimating their magnitude, and

Table 2 Characteristics of LCA research analyzed in the present study

S.No.	Biomass source	Location	System boundary	Functional unit	Software	LCIA method	Database	Year	References
1	Used rice bran oil (URBO)	India	CTGa	1 kg biodiesel production	OpenLCA	CML 2001	Ecoinvent 3.5	2022	(Pradhan et al. 2022)
2	Safflower	Turkey	CTGa	Using a new medium-sized car for 100 km	SimaPro	Eco-Indicator 99; IPCC	Ecoinvent	2022	(Isler-Kaya and Kara-osmanoglu 2022)
3	Rapeseed oil	UK	CTGa	1-ton biodiesel production	n/i	CML 2001	n/i	2022	(Gupta et al. 2022)
4	Used cooking oil	India	CTGa	1 ton of Biodiesel	SimaPro	ReCiPe 2016	Ecoinvent	2022	(Bhonsle et al. 2022)
5	Sugarcane bagasse	India	CTGa	1 kg of Bio-CNG	OpenLCA	ReCiPe 2016	n/i	2022	(Munagala et al. 2022)
6	Safflower	Malaysia	CTGa	1 MJ bioethanol produced	SimaPro	IMPACT 2002+	Ecoinvent 3.0	2022	(Hosseinzadeh-Bandbafha et al. 2022)
7	Rice straw	China	FTGa	1 metric ton dry rice straw	OpenLCA	CML 2001	Ecoinvent 3.7	2022	(Alengebawy et al. 2022)
8	Date seed oil	Oman	CTGa	1000 kg of biodiesel produced	n/i	CML-IA baseline V3.06; ReCiPe 2016	n/i	2021	(Al-Mawali et al. 2021)
9	Date palm waste	Tunisia	CTGa	1 GJ synthetic diesel and gasoline	SimaPro	ReCiPe 2016	Ecoinvent; PSILCA	2021	(ben Hnich et al. 2021)
10	Palm oil	Mexico	CTG	$1.6\times10-2$ m^3 (13.13 kg) of renewable diesel	CCalC2 V1.1	n/i	CCalC2 database	2021	(Arguelles-Arguelles et al. 2021)
11	Waste loquat seed oil	Oman	CTGa	1000 kg of biodiesel produced	SimaPro	CML-IA baseline V3.06, ReCiPe 2016	SimaPro 8.0 database	2021	(Al-Muhtaseb et al. 2021)
12	Borassus flabellifer	India	CTGa	1 kg biodiesel	SimaPro v9.0	IMPACT 2002+	Ecoinvent 3.0	2021	(Mukhopadhyay and Chakraborty 2021)
13	Crop residues	China	CTG	1t of crop residues	Gabi	CML2001	GaBi database	2021	(Yang et al. 2021)
14	Estuarine microalgae	India	CTGa	one-hectare area	OpenLCA	CML – IA	Ecoinvent 3.6	2020	(Saranya and Ramachandra 2020)
15	Spent coffee grounds	United Kingdom	CTG	treatment of 1 tonne of SCG	Gabi	ReCiPe 2016	Ecoinvent 3.3	2020	(Schmidt Rivera et al. 2020)
16	Used cooking oil	Greece	CTGa	production of 1 ton biodiesel	SimaPro 8	ReCiPe	Ecoinvent 3	2020	(Foteinis et al. 2020)
17	Pine residues; switchgrass	Southeastern United States	CTGa	1 MJ of biofuel produced	GREET	n/i	Ecoinvent; USLCA; GREET 1.8	2020	(Lan et al. 2020)
18	Rice husk	Southeast Asia	GaTGa	1 kg of biohydrogen produced	OpenLCA	CML 2001	Ecoinvent 3.5	2020	(Reaño 2020)
19	Rice straw	India	CTGa	1 L of ethanol	OpenLCA	ILCD 2011 midpoint	Ecoinvent 3.3	2020	(Sreekumar et al. 2020)
20	Forest residues	USA	CTGa	1 MJ of renewable diesel	Gabi	n/i	Ecoinvent; GaBi 9 database	2020	(Spatari et al. 2020)

Table 2 (continued)

S.No.	Biomass source	Location	System boundary	Functional unit	Software	LCIA method	Database	Year	References
21	Oleaginous yeast	India	CTGa	300 L biodiesel	SimaPro v8.0	ReCiPe End-Point (H); IMPACT 2002+	Ecoinvent 3.1	2020	(Chopra et al. 2020)
22	Desmodesmus	China	CTGa	1 kg microalgae	n/i	n/i	n/i	2020	(Li et al. 2020)
23	Waste cooking oil	Malaysia	CTGa	1000 kg of biodiesel production	SimaPro	CML 2001 and Eco-indicator 99	Agribalyse; NEEDS	2019	(Chung et al. 2019)
24	Microalgae	Taiwan	CTGa	100 MJ petrofuel produced; 1 kg biofuel produced	SimaPro	ReCiPe	Ecoinvent 3.5	2019	(Wu et al. 2019)
25	Microalgal biomass	India	CTGa	mass of CO_2 equivalent released per MJ of fuel produced	Gabi	n/i	GaBi 7 database	2019	(Banerjee et al. 2019)
26	Biowaste	EU-27	CTGa	1 MW of Biomethane	SimaPro	IMPACT 2002+	Ecoinvent 3.2	2019	(Ardolino and Arena 2019)
27	Palm oil	Malaysia	GaTGa	1 tonne of anhydrous bioethanol produced.	SimaPro	CML 2 baseline 2000	SimaPro database	2019	(Mohd YUSOF et al. 2019)
28	Jatropha	Cameroon	WTW; WTT	1 MJ of Jatropha vegetable oil	SimaPro	n/i	Ecoinvent 2.0; JEC E3-database	2019	(Vrech et al. 2019)
29	Sweet potato	Brazil	CTGa	1 MJ of bioethanol produced	SimaPro	CML IA Baseline	Ecoinvent 3.1	2018	(Costa et al. 2018)
30	Phormidium Autumnale	Brazil	CTGa	1 kg biodiesel	Chromatography Station T2100p (Plus Edition) v9.04 software	n/i	n/i	2018	(Siqueira et al. 2018)
31	Sweet potato	China	CTGa	1000 L of 99.5% bioethanol production	Microsoft Excel 2013	CML - IA 2000	CML database	2017	(Zhang et al. 2017)
32	Microalgae	Europe; Brazil	CTG	1 MJ of Biofuel	–	IMPACT 2002+	Ecoinvent; ELCD	2017	(Carneiro et al. 2017a)
33	Beef tallow	Brazil	CTGa	1 MJ of biodiesel	GaBi	CML 2001	GaBi 6.3 database	2017	(Sousa et al. 2017)
34	Pulp and paper sludge	n/i	CTGa	1 MJ of bioethanol produced	SimaPro	CML - IA	Ecoinvent 3.1; Agri-footprint	2016	(Sebastião et al. 2016)
35	Microalgae	India	WTT; TTW	1 kg biodiesel	SimaPro	ReCiPe midpoint (E)	SimaPro database	2016	(Gnansounou and Kenthorai Raman 2016)
36	Soybean; Rapeseed oil	Spain	CTGa	1 GJ	SimaPro	ReCiPe	Ecoinvent 3.1; PestLCI; MAGRAMA	2016	(Fernández-Tirado et al. 2016)
37	Rice straw	India	CTG	1 MJ of transportation fuel	Excel	n/i	NREL reports; Lab experiments	2016	(Soam et al. 2016)

Table 2 (continued)

S.No.	Biomass source	Location	System boundary	Functional unit	Software	LCIA method	Database	Year	References
38	Lignocellulosic	UK	CTG	1 L of ethanol	GaBi	CML 2	Ecoinvent 2.2	2015	(Jeswani et al. 2015)
39	Palm oil	n/i	CTG	n/i	SimaPro	CML 2000	Ecoinvent 2.1	2015	(Arpornpong et al. 2015)
40	Sugarcane	Brazil	WTT; TTW	1 ton of sugarcane; 1 km of car operation	SimaPro	ReCiPe midpoint (E)	Ecoinvent 3.0	2015	(Gnansounou et al. 2015)
41	Rice husk	India	CTGa	1 L of ethanol	–	–	NREL	2015	(Mandade et al. 2015)
42	Wheat; Rapeseed	Spain	WTT	1 MJ biofuel energy content	SimaPro	CML 2000	Ecoinvent 2.0	2015	(Gallejones et al. 2015)
43	Sugarcane; Soybean; and Palm oil	Brazil	WTT	1 MJ of energy released by combustion of bioethanol, biodiesel	SimaPro	CML 2 Baseline 2000	n/i	2014	(Rocha et al. 2014)
44	Microalgae	US	CTGa	2 million gallons of microalgal biodiesel per year	n/i	TRACI 2	Ecoinvent 2.0; LCA Food DK; EIA; DOE; ERS; USDA	2014	(Soratana et al. 2014)
45	Palm oil; Jatropha; Rapeseed oil	n/i	CTGa	1 MJ of produced renewable diesel	GaBi	n/i	GaBi database	2014	(Uusitalo et al. 2014)
46	Maize; Sugar beet; Sugarcane; Wheat	n/i	CTGa	1 kg of ethanol	SimaPro	ReCiPe	SimaPro database	2014	(Muñoz et al. 2014)
47	Palm oil	Malaysia	WTW	1 kg of palm oil methyl ester (PME)	GaBi	Eco-Indicator 99	Ecoinvent	2013	(Lee and Ofori-Boateng 2013)
48	Sweet potato	China	CTGa	1000 L of bioethanol produced	Microsoft Excel 2003	CML-IA	n/i	2013	(Wang et al. 2013b)
49	Rapeseed oil	Spain	CTGa	1 ton of biofuel	SimaPro	CML	Ecoinvent	2013	(González-García et al. 2013a)
50	Wheat straw	UK	WTW	to drive 1 km in a Flexible-fuel vehicle (FFV)	SimaPro	CML Baseline 2000	Ecoinvent 2.2	2013	(Wang et al. 2013a)
51	Jatropha	India	CTG	1 ton of biodiesel produced	n/i	n/i	CEA database; IPCC	2012	(Kumar et al. 2012)
52	Wheat straw	n/i	WTGa	1 km distance by a passenger car; 1 kg of ethanol produced	SimaPro	ReCiPe midpoint	Ecoinvent	2012	(Borrion et al. 2012)
53	Microalgae	n/i	WTW	8.94 10^{10} MJ/year of biodiesel	GREET	IMPACT 2002+; TRACI	ETH-ESU; Ecoinvent 2.0; BUWAL250; IDEMAT	2012	(Soratana et al. 2012)

Cradle to Gate (CTGa), Field to Gate (FTGa), Farm to Gate (FaTGa), Cradle to Grave (CTG), Gate to Gate (GaTGa), Well to Wheel (WTW), Well to Tank (WTT), Tank to Wheel (TTW), Well to Gate (WTGa), n/i: not included

propagating them through the LCA model to determine their impact on the final results.

The LCA performs a sensitivity analysis by varying a parameter's upper and lower limits and running simulations with different combinations of parameter values. This method is used along with scenario analysis to evaluate how uncertain factors may affect the results (Eggemann et al. 2020). Sensitivity analysis determines which step in the biofuel manufacturing life cycle is directly responsible for the negative environmental impacts.

In LCA, a scenario refers to a specific situation or future projection that is represented through a defined set of assumptions and choices. These scenarios encompass the system boundaries and assumptions about the inputs and outputs of the analyzed product or system (Arguelles-Arguelles et al. 2021). They serve as a means to capture distinct conditions and provide a foundation for analysis and comparison. Scenario analysis entails evaluating the environmental impacts under various hypothetical scenarios or alternative future pathways (Yang et al. 2021). It allows for the exploration of different configurations, technologies, materials, or strategies within the system to gain insights into their potential effects on the overall environmental performance. As shown in Fig. 5, the majority of the reviewed studies (64%) have considered scenario-based analysis. By comparing different scenarios, decision-makers can assess the consequences of different choices and make well-informed decisions.

Interpretation

The interpretation phase considers the impact assessment results and the inventory analysis outcomes (ISO 2006a). The analysis results, together with all the decisions and assumptions made during the process, are examined and analyzed, and broad conclusions are drawn (Osman et al. 2021b). The aim is to arrive at conclusions that back up a decision or produce an LCA result that is easy to comprehend.

Environmental sustainability of biofuels

Carbon footprint

Most LCA biofuel studies focus on GHG emissions and reductions compared to fossil fuels. Biofuel LCA studies also consider other environmental impact categories, such as acidification, eutrophication, photochemical smog, human toxicity, and ecotoxicity. LCA studies show inconsistent conclusions for Global Warming Potential (GWP), ranging from favorable to unfavorable, even for the same feedstock. This is due to assumptions, data sources, allocation procedures, and LUC variations. GHG emissions are converted to g CO_2 eq. per MJ to compare data from different studies. The lower heating values are assumed to be 27 MJ/kg and 38 MJ/kg for bioethanol and biodiesel, respectively (Yan and Crookes 2009), whenever these statistics are not stated in the original research. The use phase of the biofuel is not considered in the subsequent analysis.

First-generation biofuels

The estimated GHG emission of first-generation bioethanol ranges from 0.006 to 167 g CO_2 eq. / MJ, and that for biodiesel varies from − 7.3 to 329 g CO_2 eq. / MJ, as shown in Fig. 6a and b, respectively. When the average values are taken into account, the GWP of most of the first-generation biofuels is less than those of conventional gasoline (72.8 to 96 g CO_2 eq. / MJ (González-García et al. 2010a, b; Ou et al. 2009) and diesel (80 to 120 g CO_2 eq. / MJ (Ou et al. 2009; Shirvani et al. 2011; Passell et al. 2013), except for Rapeseed biodiesel.

Some LCA studies that have been analyzed show rapeseed biodiesel emits 2.8 to 350 g CO_2 eq. / MJ of energy (González-García et al. 2013b; Uusitalo et al. 2014; Gupta et al. 2022). Compared to fertilizer application in the agricultural stage (which consumes 70% more energy and 80% more GWP) (Gallejones et al. 2015), the contribution from the phases of drying and cooling of seeds, oil extraction, and refining was minimal. According to the study by (Arguelles-Arguelles et al. 2021), GHG emission was negative for biodiesel production from palm oil because of the amount of carbon dioxide captured during the cultivation phase being more than biofuel's GHG emission. Most studies on palm oil, soybeans, and sugarcane (Lee and Ofori-Boateng 2013; Rocha et al. 2014; Arpornpong et al. 2015; Fernández-Tirado et al. 2016; Munagala et al. 2022) indicated lower GHG emissions than conventional gasoline. The growing and harvesting phase (Phase 1, Fig. 2a) of biofuel LCA is critical for biofuel generation. As a result, the quantity of fertilizer, pesticides, and fuel used during the agricultural phase plays a critical role in determining the outcomes and environmental effects of the biofuel, as shown in Table 3. Most reviewed studies show a reduction in GHG emissions for first-generation biofuel compared to conventional diesel and gasoline.

Second-generation biofuels

The GHG emissions of second-generation biofuels vary from − 15.4 to 178.7 g CO_2 eq. / MJ for bioethanol and − 0.21 to 113.8 g CO_2 eq. / MJ for biodiesel, as shown in Fig. 6c and d, respectively. These discrepancies result from various factors such as the feedstock used, the production routes, different technological assumptions, and differences in methodology

Table 3 Stagewise GHG contribution of biofuel production

Biomass source	Percentage of GHG Contribution				
	Cultivation	Transportation	Oil / Juice processing	Biofuel production	References
Jatropha	55–62	8–10	28–32	5–8	(Kumar et al. 2012; Kalaivani et al. 2014; Vrech et al. 2019; Uppalapati et al. 2022)
Soybean	34–40	16–23	18–24	18–26	(Rocha et al. 2014; Fernández-Tirado et al. 2016; Uppalapati et al. 2022)
Calophyllum inophyllum	40	19	21	20	(Uppalapati et al. 2022)
Rapeseed	68–72	8–11	5–7	11	(Uusitalo et al. 2014; Fernández-Tirado et al. 2016; Gupta et al. 2022)
Sugarcane	53–63	6–10	12–15	20–21	(Rocha et al. 2014; Gnansounou et al. 2015; RSB 2020; Munagala et al. 2022)
Palm Oil	40	6–12	7–9	41–46	(Lee and Ofori-Boateng 2013; Uusitalo et al. 2014; Mohd YUSOF et al. 2019)
Microalgae	30–33	–	20–23	43–47	(Pragya and Pandey 2016; Deshmukh et al. 2019; Uppalapati et al. 2022)
Oleaginous yeast	40–43	–	2–6	49	(Jain et al. 2018; Chopra et al. 2020)

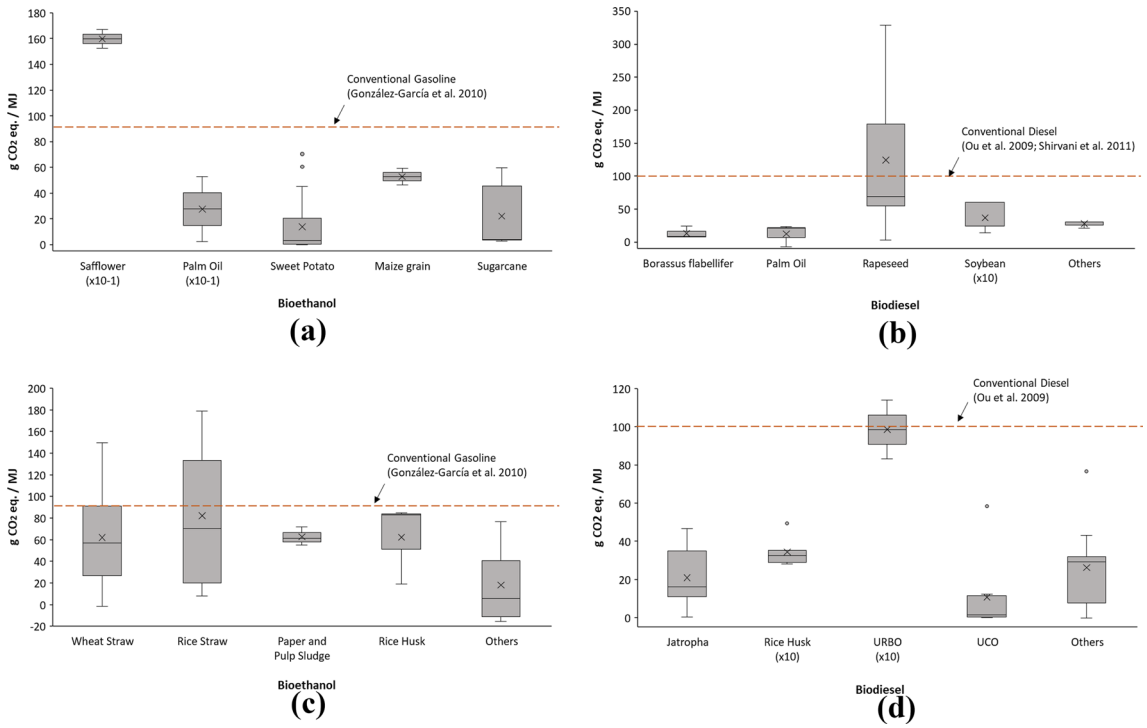


Fig. 6 GHG emissions for **a** bioethanol and **b** biodiesel from first-generation biofuels; **c** bioethanol and **d** biodiesel from second-generation biofuels

used for the analysis. It should be highlighted that technological uncertainties are particularly important when evaluating their potential because advanced biofuels have not yet reached full commercialization. As a result, the quality of the information that is now accessible is not as strong as in the case of the well-known first-generation biofuels (Mohd YUSOF et al. 2019; Kovacs et al. 2022). The GWP of lignocellulosic biofuel produced from agricultural and forestry wastes is often lower than biofuel produced from energy crops (Gandam et al. 2022). This is mainly caused by N_2O emissions from fertilizers while growing energy crops (Olukoya et al. 2014; Jeswani et al. 2020). Assuming that

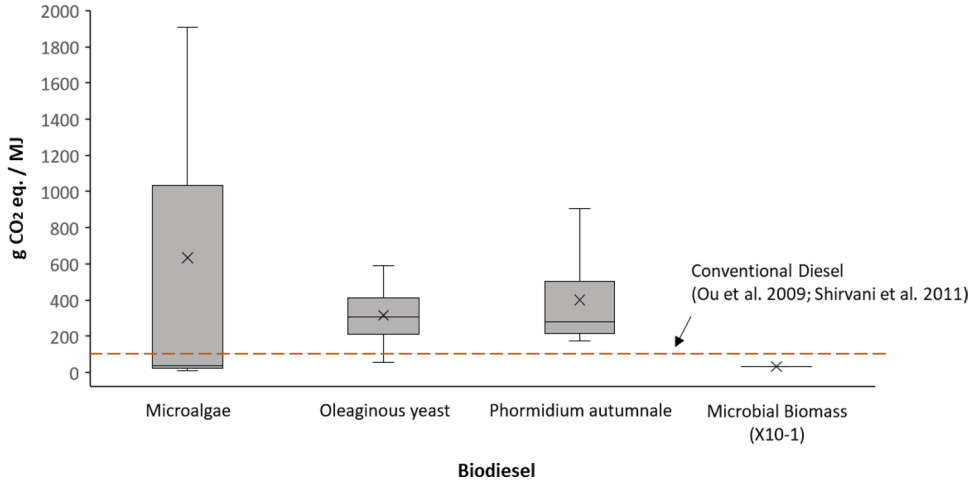


Fig. 7 GHG emissions for biodiesel from third-generation biofuels

residues do not have any environmental effects and are solely linked to the primary crop from which the waste is generated, then the adverse effects related to the primary crop can be prevented by using the residues instead. The production of lignocellulosic bioethanol involves the assumption that the leftover lignin will generate heat and power to meet the energy needs of the process. Any excess electricity produced is then supplied to the grid (Sawatdeenarunat et al. 2015). This results in the biofuel production system being acknowledged for preventing GHG emissions from a specific amount of grid power (Jiang et al. 2022). In some instances, such as with feedstocks like spent coffee grounds, forest residue, and rice straw, the benefits from generating power and other byproducts outweigh the emissions from producing biofuels.

Some of the reviewed studies show negative GHG emissions for the production of biofuels.(Jeswani et al. 2015) analyzed the LCA of ethanol production from lignocellulosic biomass and compared the bioethanol production from wheat straw, miscanthus, poplar, and forest residue. The analysis shows that the GWP of bioethanol production is negative for all four feedstocks, mainly due to the production of co-products. The study by(Schmidt Rivera et al. 2020) compares the different valorization routes of spent coffee grounds for the production of value-added products such as biodiesel. The method of incineration was found to be the most sustainable option because it had a total of 14 net negative effects on the environment, including the emission of greenhouse gases (−0.212 g CO_2 eq. per MJ). The high energy output of waste-to-energy conversion technologies and credits for replacing the electricity mix that is dominated by fossil fuels mitigate the negative impacts of these valorization routes.(Pradhan et al. 2022) analyzed biodiesel production from used rice bran oil through esterification and transesterification processes. The GHG emissions were found to be very high (831.5 to 1138.4 g CO_2 eq. / MJ) than conventional diesel (102 g CO_2 eq. / MJ (Ou et al. 2009), which has similar results as the study by (Sun et al. 2022). However, the analysis shows that the vanadium-impregnated solid catalyst results in a 26.96% reduction in GWP compared to a commercial hydroxyapatite-supported catalyst, shown in Fig. 6d.

Third-generation biofuels

Among the studies that were evaluated, only 7 LCA studies were found to have calculated the GHG emissions for third-generation biofuels. However, these studies used different approaches, methods, and assumptions when it came to managing feedstocks, nutrients, and co-products, as well as setting system boundaries and process designs. Consequently, the GHG emissions have varied significantly, as indicated in Fig. 7, ranging from 10.2 to 1910 g CO_2 eq. / MJ. These findings suggest that, depending on the assumptions, microalgae diesel can significantly reduce or increase GHG emissions compared to diesel (Soratana et al. 2014). Despite being in its present development stage, many research works have concluded that algal biodiesel generates more GHGs throughout its lifespan than fossil fuels (Banerjee et al. 2019). This is majorly due to the low algal yield and the substantial energy consumption during the cultivation, harvesting, and drying stages (Mu et al. 2014; Banerjee et al. 2019).

A study by (Saranya and Ramachandra 2020) evaluated the LCA of estuarine microalgae for biodiesel production. The analysis shows that the scenario which uses wastewater as a nutrient has less GHG emissions of 0.85 kg CO_2 eq. per kg of biodiesel (22.4 g CO_2 eq. / MJ) than other scenarios. There was a 67.9 to 85.4% reduction in GHG emissions when microalgal biodiesel was used instead of fossil diesel. However, the study by (Soratana et al. 2012) shows that none of the four different conditions considered for analysis for biodiesel production meets the Renewable Fuel Standard, which requires a 50% reduction in GHG emissions for advanced biofuels. The analysis shows greenhouse gas

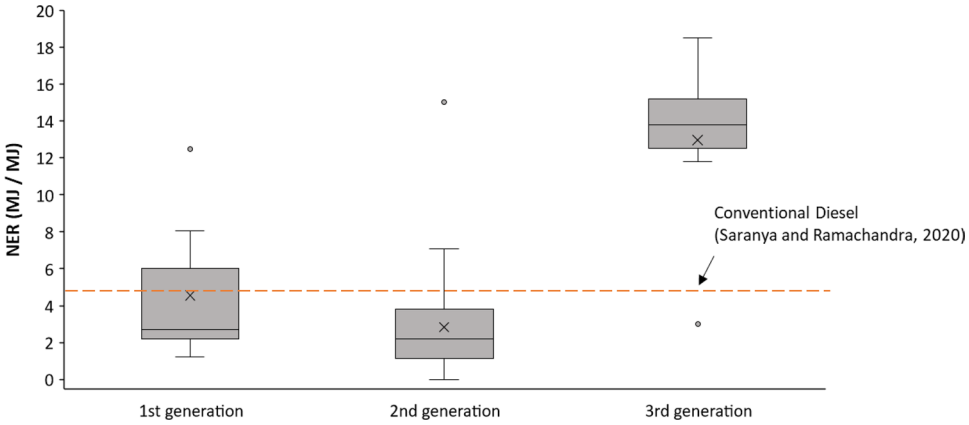


Fig. 8 NER for first, second, and third-generation biofuel

emissions range from 854 to 1910 g CO_2 eq. / MJ. Another research by (Chopra et al. 2020) examined the environmental impact of producing biodiesel and bio-crude from oleaginous yeast and found GHG emissions ranging from 56 to 600 g CO_2 eq. / MJ. Therefore, more investigation is needed to ensure the sustainability of the process.

Energy Use

Several indicators, including fossil energy consumption, primary, secondary, or cumulative energy demand, and net energy ratio (NER), are employed in LCA studies to assess energy usage in the life cycle of biofuels (Arvidsson et al. 2012). However, considering that lowering dependency on fossil fuels and ensuring energy security are important goals of national biofuel policies, in addition to mitigating climate change, many concentrated on fossil energy consumption. The NER ratio measures how much energy the biofuel produces in relation to total energy intake. If the NER is less than 1, the system is not advantageous since more energy is needed to produce the biofuel than it contains (Carneiro et al. 2017b). As shown in Fig. 8, most research anticipates that fossil energy use for 1st and 2nd generation biofuels varies across different biofuels ranging from 1.23 to 12.49 MJ / MJ for 1st generation and from 0.003 to 15.04 MJ / MJ for 2nd generation biofuel. Different farming techniques, feedstock production, conversion technology, and allocation strategies are some of the causes of these discrepancies. The assumptions about the type of energy employed in the conversion process also impact the outcomes. The energy demand of LCA of microalgae biofuels is considerably more diverse, ranging from 3 to 18.5 MJ / MJ. The reasons for this variation include technological uncertainty and the range of prospective feedstocks and manufacturing systems.

(Al-Mawali et al. 2021) estimate a NER of 2.17 MJ / MJ for the generation of biodiesel from waste date seed oil, which is comparable to the energy ratio of 3.15 MJ / MJ reported by (Pleanjai and Gheewala 2009) for palm oil biodiesel. (Al-Muhtaseb et al. 2021) also reported a similar NER (2.23 MJ/ MJ) for waste loquat seed oil biodiesel with an enhanced bifunctional catalytic system. (Uppalapati et al. 2022) estimated the NER for different feedstocks for biodiesel production. The analysis shows that Jatropha biodiesel has the highest NER (15 MJ / MJ), which is in line with the study by (Abbas et al. 2016) and (Silitonga et al. 2016) and the lowest for microalgae (3 MJ / MJ). Similar NER was reported by (Pragya and Pandey 2016), (Ajayebi et al. 2013), and (Firoz 2008) for microalgae (2.89 MJ / MJ) and Soybean (12.49 MJ / MJ). (Reaño 2020) and (Sreekumar et al. 2020) analyzed biofuel production from rice husk and straw. The EROI (energy return on investment), similar to NER, ranges from 0.6 to 1.59 MJ / MJ. This shows that the process is net energy positive and hence desirable. Similarly, (Mandade et al. 2015) analyze the lignocellulosic feedstock for ethanol production, and EROI values vary from 2 to 7.1 for different allocation methods.

Water use

Water consumption can be significant in producing bio-feedstocks, especially for first-generation biofuels (Dominguez-Faus et al. 2009). The use of irrigation water for certain feedstocks could pose a problem as it may compete with the water requirements for other purposes like producing food (Gerbens-Leenes et al. 2009). The total amount of water consumed worldwide might drastically increase due to increasing agricultural biomass production for biofuels (Hammond and Li 2016). In regions currently experiencing water scarcity, this extra water demand could significantly worsen the overall environmental effects of biofuels. As shown in Fig. 9, the water consumption of biofuel production ranges from 36.2 to 540 m^3 / GJ for first-generation biofuel and from 1.28 to 188.8 m^3 / GJ for second-generation biofuel. When scarcity of water in a particular region is considered, not just the amount of water used, the water footprints (WF) of biofuels can be

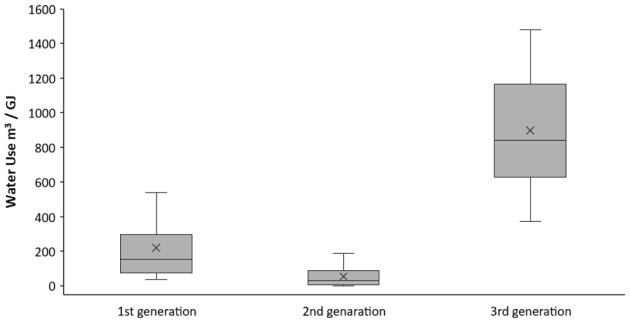


Fig. 9 Water consumption for biofuel production based on reviewed studies

50 to 240 times larger than those of fossil fuels (Berger et al. 2015). This is due to a significant portion of water use in biofuel production in water-stressed regions. In this analysis, the cultivation stage is not included for the second-generation biofuel. It is based on the assumption that the cultivation of feedstock has already been accounted for the LCA of primary food crops or other agricultural products, as they are typically byproducts or residues from those processes (Soam et al. 2016).

(Mandade et al. 2015) analyzed the water life cycle for Indian agro-industrial lignocellulosic biomass-based biofuel production. The analysis shows sorghum stalk biofuel to be more water efficient and sugarcane to be least for biofuel production as it consumes more water for Phase 1 (Fig. 2a) system boundary of LCA of biofuel. (Mathioudakis et al. 2017) analyzed the water footprint of agricultural residues and showed the WF of rice straw to be 129 L / kg. (Xie et al. 2017) analyzed for the ethanol production from cassava and sorghum and reported a WF ranging between 1760 and 5290 L per liter of ethanol, mostly agricultural, contributes towards the WF. Similarly, (Arguelles-Arguelles et al. 2021) show that more than 90% of the water used to produce biodiesel from palm oil ranges from 218 to 540 m^3 / GJ was for agriculture of palm plantation. Similar results were shown by (Gerbens-Leenes and Hoekstra 2010).

As shown in Fig. 9, the WF for third-generation biofuel is the highest ranging from 373.4 to 1481 m^3 / GJ (Siqueira et al. 2018; Wu et al. 2019). The water footprint of biofuel production from microalgae can vary depending on production methods and water availability in the production region (Banerjee et al. 2019). The WF considers the process water, material inputs for every stage of LCA, and water credit correlated with co-products (Mekonnen and Hoekstra 2011; Siqueira et al. 2018). The total WF comprises both the blue and green WFs, where the blue WF reflects the amount of water needed from local sources such as surface or groundwater for irrigation purposes, while the green WF represents the quantity of rainwater and soil moisture utilized by plants (Mekonnen and Hoekstra 2011).

Land use change

Land-use change (LUC) either directly or indirectly causes GHG emissions. Since the biomass that powers these systems is produced on the land, land use is a crucial factor in the research of bioenergy systems. LUC is a significant source of GHG emissions, contributing 13 Gt CO_2 – eq globally in 2019 (IPCC 2022). Only 46% of the reviewed studies considered land-use change, as shown in Fig. 3. The lack of attention on advanced LCA studies, such as land-use change, was caused by the unavailability of a significant land-use database (Harris et al. 2015). Based on research conducted by (Fargione et al. 2008; Lapola et al. 2010), it has been found that the amount of GHG emissions resulting from direct and indirect LUC can be considerably higher than the GHG reductions achieved by replacing fossil fuels with biofuels.

Studies that include LUC tend to report a higher total environmental impact for biofuel production compared to studies that exclude LUC. This is due to the additional emissions and ecological impacts associated with land conversion. A study by (Isler-Kaya and Karaosmanoglu 2022) reported land use impact on biodiesel production from safflower to be 9.21 and 9.94 PDF*m^2*yr ("potential disappeared fraction of species over a certain area over a certain period") for safflower oil methyl ester (SOME) and ethyl ester (SOEE), which are similar as the study by (Spinelli et al. 2013). Study by (Humpenöder et al. 2013) shows that there was a 50% reduction in GHG emissions without taking LUC into account compared to fossil fuel. (Bhonsle et al. 2022) analyzed biodiesel production from UCO. The analysis shows land use midpoint category was highest for oil boiler use (86.6 m2a crop eq.) for conventional biodiesel and UCO use (6.59 m2a crop eq.) for room temperature scenario. (Chung et al. 2019) also analyzed for the used cooking oil and showed land use impact of 4.72E-01 PDF*m^2*yr for catalyst preparation, 4.33E-01 PDF*m^2*yr for pretreatment, and 1.26 PDF*m^2*yr for the transesterification process. Similar studies were done by (Fernández-Tirado et al. 2016; Gnansounou and Kenthorai Raman 2016; Carneiro et al. 2017a; Vrech et al. 2019). While biofuels can offer advantages in reducing fossil fuel use and associated emissions, the inclusion of LUC-related impacts may diminish these advantages, especially if significant land conversion occurs.

Other environmental impacts

Besides carbon footprint, the environmental impact categories examined in biofuel LCA studies are acidification,

Table 4 Comparison of damage categories using Eco-Indicator 99 and ReCiPe 2016 LCIA method

Feedstocks	Unit	Human health		Ecosystem quality		Resources	
		Eco-Indicator 99	ReCiPe 2016	Eco-Indicator 99	ReCiPe 2016	Eco-Indicator 99	ReCiPe 2016
Rapeseed	%	47.6	94.5	5.7	3.2	46.6	2.4
Palm Oil	%	54.8	91.9	21.3	7.9	23.9	0.2
Used Cooking Oil	%	31.8	92.4	2.6	3.1	65.6	4.6
Jatropha	%	44.2	93.9	5.1	3.6	50.8	2.5
Microalgae	%	56.9	94.9	2.2	3.5	40.9	1.7

eutrophication, and ecotoxicity. About 60% of the reviewed studies analyzed more comprehensive impact categories. Since the reviewed LCA studies employed various methodologies to assess other environmental impacts, comparing them and providing a meaningful range of impacts for different biofuels is difficult (Kalaivani et al. 2014). Furthermore, there is variability in the scope of the studies reviewed. Some studies focus primarily on assessing the environmental impacts of biofuels within a cradle-to-gate system boundary, which considers the stages from raw material extraction to the point of leaving the production facility. On the other hand, certain studies adopt a broader cradle-to-grave system boundary, encompassing the entire life cycle of the biofuel, including additional stages such as use, maintenance, and end-of-life considerations. This variation in system boundaries leads to differences in the extent and depth of the assessments conducted, affecting the comprehensiveness and breadth of the environmental impact analysis (Jolliet et al. 2003). The results of the research are also impacted by the assumptions made regarding the type of vehicle that biofuels are used in. (Arpornpong et al. 2015). Some studies indicate that reducing GHG emissions by using biofuels instead of fossil fuels can have negative consequences on the environment, such as increasing acidification, eutrophication, respiratory issues caused by organic and inorganic pollutants, and ecotoxicity (Borrion et al. 2012; Arguelles-Arguelles et al. 2021; Yang et al. 2021). Figure 10a shows the impact categories of biofuel production from different sources using the IMPACT 2002 + impact characterization. Impact characterization aids in the identification of primary drivers of environmental impacts throughout a life cycle. Through the comparison of impact scores assigned to various inputs or outputs, it allows for the identification of hotspots, which refer to specific stages, processes, or materials that exert a substantial influence on the overall environmental performance (Isler-Kaya and Karaosmanoglu 2022).

Ecotoxicity potential (EP) is mainly attributed to fertilizers and pesticides used in Phase 1 of the LCA system boundary (Fig. 2a). (Lee and Ofori-Boateng 2013; Gallejones et al. 2015) showed similar results for rapeseed biofuel and palm oil, where the use of fertilizers caused 42% and 50% of total EP. However, the midpoint impacts, such as non-renewable energy and respiratory inorganics, which contributes to human health and resource depletion (González-García et al. 2013b), have the highest single score for biofuel production, shown in Fig. 10b. According to the research conducted by (Kalaivani et al. 2014; Patel and Singh 2023), the ecotoxicity of jatropha biodiesel ranges from 35 to 160 kg TEG (triethylene glycol) per kg of biodiesel. The study found that phosphorus fertilizer accounts for 45% of the total EP, making it the largest contributor, followed by urea, which contributes 20% of the EP. Table 4 compares the endpoint or damage categories of Eco-indicator 99 and ReCiPe 2016 impact method for biofuel production. By comparing the two methods, it helps to assess the consistency and reliability of the results obtained. It allows researchers and practitioners to identify any discrepancies or variations in impact assessment outcomes and evaluate the methodological robustness of each approach. enables informed decision-making by considering different perspectives and approaches to impact assessment. Table 5 shows the environmental impact categories on a relative scale.

Socio-Economic impacts of biofuel production

The socio-economic impacts of biofuels refer to the effects that the production, use, and promotion of biofuels have on society and the economy. These impacts can be positive and negative and vary depending on factors such as the specific biofuel feedstock, production methods, local contexts, and policy frameworks (Kim and Dale 2008). Biofuel production can create employment opportunities across the value chain, including farming, harvesting, processing, transportation, and distribution. This can benefit rural communities and contribute to local economic development. Furthermore, biofuels reduce dependence on fossil fuels, promoting energy security by diversifying energy sources (Cambero and Sowlati 2014; Yang et al. 2015). This can enhance a country's energy independence and reduce vulnerability to price fluctuations and supply disruptions in the global oil market. However, the cultivation of biofuel feedstocks may compete with food production, potentially leading to

increased food prices and reduced food security (Collotta et al. 2019). This can particularly affect regions where land resources are limited or where the agricultural sector is already under stress.

Compared to environmental impacts, economic and social indicators are rarely included in the LCA. Few studies included the socio-economic aspect of biofuel (van Dam et al. 2009; Stephenson et al. 2010; Ekener-Petersen et al. 2014; Ekener et al. 2018; Collotta et al. 2019). Key factors highlighted in these studies encompass indicators such as transportation expenses, food costs, gender equality, health-related diseases, instances of forced labor, and minimum wages., which exert a substantial impact on the broader socio-economic sustainability of biofuel production (van Dam et al. 2009; Yang et al. 2015).

Conclusion

The LCA is a complex technique that is often utilized to compute the GWP and other environmental impacts associated with biofuels. Despite this, it is frequently seen as a tool capable of providing a conclusive response to numerous concerns. This study indicates that there are no concrete solutions, and the conclusions drawn from LCA studies largely depend on factors like the study's goal and scope, the assumptions made, the methodology used, and the tools employed. Therefore, it is essential to acknowledge this and interpret the results accordingly. Furthermore, certain strategies for communicating complexity to policymakers and the genpotential disappeared fraction of specieral public are required. This analysis shows that the estimates between the studies differ significantly due to numerous LCA methodological options and uncertainties. Despite this, the available evidence is instructive. To keep up with the latest trends in evaluating the environmental effects linked to biofuel production, we critically reviewed 53 LCA studies published between 2012 and 2022, carefully examining the methodology and conclusions. Our findings suggest that second-generation biofuels have a greater potential to decrease GHG emissions than first-generation biofuels. Third-generation algal biofuels show high GHG emissions than conventional fuels at the current stage of development since their electricity consumption is high. The study critically reviewed the energy use and water use in the production of biofuels, along with several midpoint impact categories, such as acidification and eutrophication associated with biofuel production. Based on the bibliometric mapping from the WoS, we have identified that bioenergy, life cycle assessment, biofuels, and sustainability are more researched areas.

To improve the usability of LCA as an evidence-based tool for assessing the environmental sustainability of biofuels, both attributional and consequential LCA methodology and practical applications must be improved. Uncertainty and sensitivity assessments should be included in all LCA investigations to increase the reliability of the results and improve confidence in making decisions based on LCA studies. Biofuels are not independent entities but a component of much larger systems, such as those related to energy, agriculture, and forestry. Biofuels affect several ecosystem services,

Table 5 Midpoint single score based on IMPACT 2002 + LCIA method for biodiesel production

Impact category	Unit	Jatropha biodiesel (Uusitalo et al. 2014; Vrech et al. 2019)	Microalgae (Soratana et al. 2012, 2014; Siqueira et al. 2018)	Palm oil (Lee and Ofori-Boateng 2013; Arguelles-Arguelles et al. 2021)	Rapeseed biodiesel (González-García et al. 2013b; Gallejones et al. 2015)	Used cooking oil (Uusitalo et al. 2014; Bhonsle et al. 2022)
Total	%	100	100	100	100	100
Carcinogens	%	4.63	2.00	3.29	4.50	6.69
Non-carcinogens	%	2.66	1.51	2.92	2.89	1.69
Respiratory inorganics	%	22.65	31.21	34.61	25.62	23.38
Ionizing radiation	%	0.04	0.08	0.08	0.05	0.11
Ozone layer depletion	%	0.00	0.00	0.00	0.00	0.02
Respiratory organics	%	0.02	0.01	0.02	0.03	0.04
Aquatic ecotoxicity	%	0.11	0.10	0.15	0.13	1.21
Terrestrial ecotoxicity	%	3.74	3.23	5.51	4.56	2.40
Terrestrial acid/nutri	%	0.32	0.23	0.41	0.39	0.29
Land occupation	%	0.39	0.16	-0.24	0.40	-10.30
Non-renewable energy	%	36.91	32.09	26.61	35.21	50.43
Mineral extraction	%	0.06	0.01	0.07	0.08	0.05

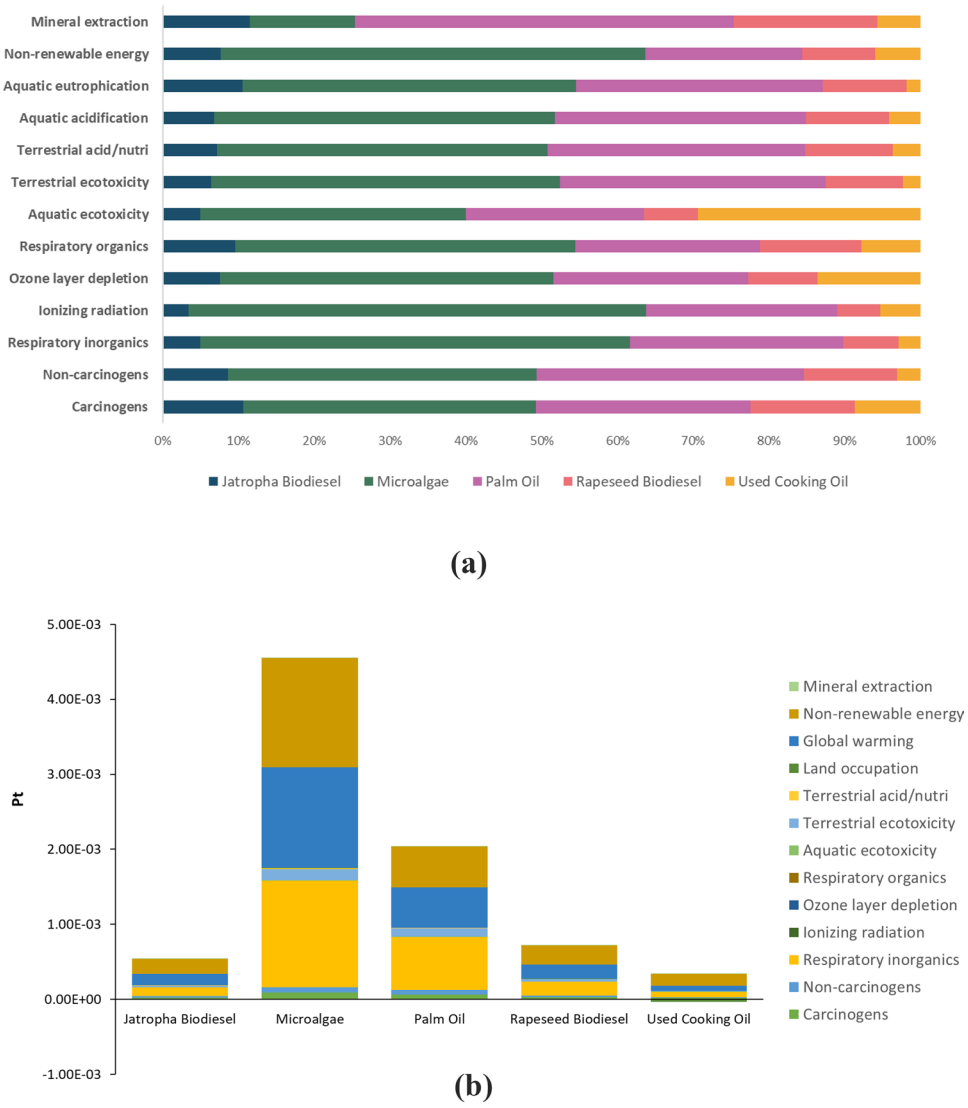


Fig. 10 Midpoint impact categories **a** Characterization score and **b** Single score of biofuel production

including food, water, and land. To prevent biofuels from being disadvantaged compared to other industries or from having advancements in this industry reversed by unsustainable practices in other industries, it is crucial to plan future policy from an integrated systems perspective.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10098-023-02596-y.

Author contributions KP: Conceptualization; Formal analysis and investigation; Writing - original draft preparation, S.K.S: Writing - review and editing; Supervision.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abbas J, Hussain S, Iqbal MJ et al (2016) Oil industry waste: a potential feedstock for biodiesel production. Environ Technol 37:2082–2087. https://doi.org/10.1080/09593330.2016.1141997

Ajayebi A, Gnansounou E, Kenthorai Raman J (2013) Comparative life cycle assessment of biodiesel from algae and jatropha: a case study of India. Bioresour Technol 150:429–437. https://doi.org/10.1016/j.biortech.2013.09.118

Al-Mawali KS, Osman AI, Al-Muhtaseb AH et al (2021) Life cycle assessment of biodiesel production utilising waste date seed oil and a novel magnetic catalyst: a circular bioeconomy approach.

Renew Energy 170:832–846. https://doi.org/10.1016/j.renene.2021.02.027
Al-Muhtaseb AH, Osman AI, Murphin Kumar PS et al (2021) Circular economy approach of enhanced bifunctional catalytic system of CaO/CeO2 for biodiesel production from waste loquat seed oil with life cycle assessment study. Energy Convers Manag 236:114040. https://doi.org/10.1016/j.enconman.2021.114040
Alengebawy A, Mohamed BA, Ran Y et al (2022) A comparative environmental life cycle assessment of rice straw-based bioenergy projects in China. Environ Res 212:113404. https://doi.org/10.1016/j.envres.2022.113404
Ardolino F, Arena U (2019) Biowaste-to-Biomethane: an LCA study on biogas and syngas roads. Waste Manag 87:441–453. https://doi.org/10.1016/j.wasman.2019.02.030
Arguelles-Arguelles A, Amezcua-Allieri MA, Ramírez-Verduzco LF (2021) Life Cycle Assessment of Green Diesel production by Hydrodeoxygenation of Palm Oil. Front Energy Res. https://doi.org/10.3389/fenrg.2021.690725
Arpornpong N, Sabatini DA, Khaodhiar S, Charoensaeng A (2015) Life cycle assessment of palm oil microemulsion-based biofuel. Int J Life Cycle Assess 20:913–926. https://doi.org/10.1007/s11367-015-0888-5
Arvidsson R, Fransson K, Fröling M et al (2012) Energy use indicators in energy and life cycle assessments of biofuels: review and recommendations. J Clean Prod 31:54–61. https://doi.org/10.1016/j.jclepro.2012.03.001
Balajii M, Niju S (2019) Biochar-derived heterogeneous catalysts for biodiesel production. Environ Chem Lett 17:1447–1469. https://doi.org/10.1007/s10311-019-00885-x
Banerjee S, Rout S, Banerjee S et al (2019) Fe2O3 nanocatalyst aided transesterification for biodiesel production from lipid-intact wet microalgal biomass: a biorefinery approach. Energy Convers Manag 195:844–853. https://doi.org/10.1016/j.enconman.2019.05.060
Bare JC, Hofstetter P, Pennington DW, de Haes HAU (2000) Midpoints versus endpoints: the sacrifices and benefits. Int J Life Cycle Assess 5:319–326. https://doi.org/10.1007/BF02978665
ben Hnich K, Martín-Gamboa M, Khila Z et al (2021) Life cycle sustainability assessment of synthetic fuels from date palm waste. Sci Total Environ 796:148961. https://doi.org/10.1016/j.scitotenv.2021.148961
Berger M, Pfister S, Bach V, Finkbeiner M (2015) Saving the planet's climate or Water Resources? The Trade-Off between Carbon and Water Footprints of European Biofuels. Sustainability 7:6665–6683. https://doi.org/10.3390/su7066665
Bharti MK, Chalia S, Thakur P et al (2021) Nanoferrites heterogeneous catalysts for biodiesel production from soybean and canola oil: a review. Environ Chem Lett 19:3727–3746. https://doi.org/10.1007/s10311-021-01247-2
Bhonsle AK, Singh J, Trivedi J, Atray N (2022) Comparative LCA studies of biodiesel produced from used cooking oil using conventional and novel room temperature processes. Bioresour Technol Rep 18:101072. https://doi.org/10.1016/j.biteb.2022.101072
Borrion AL, McManus MC, Hammond GP (2012) Environmental life cycle assessment of bioethanol production from wheat straw. Biomass Bioenergy 47:9–19. https://doi.org/10.1016/j.biombioe.2012.10.017
BRDI (2008) Increasing feedstock production for Biofuels Economic Drivers, Environmental Implications, and the role of Research. The Biomass Research and Development Board (US)
Cambero C, Sowlati T (2014) Assessment and optimization of forest biomass supply chains from economic, social and environmental perspectives – A review of literature. Renew Sustain Energy Rev 36:62–73. https://doi.org/10.1016/j.rser.2014.04.041
Carneiro MLNM, Pradelle F, Braga SL et al (2017a) Potential of biofuels from algae: comparison with fossil fuels, ethanol and biodiesel in Europe and Brazil through life cycle assessment (LCA). Renew Sustain Energy Rev 73:632–653. https://doi.org/10.1016/j.rser.2017.01.152
Carneiro MLNM, Pradelle F, Braga SL et al (2017b) Potential of biofuels from algae: comparison with fossil fuels, ethanol and biodiesel in Europe and Brazil through life cycle assessment (LCA). Renew Sustain Energy Rev 73:632–653. https://doi.org/10.1016/j.rser.2017.01.152
Chamkalani A, Zendehboudi S, Rezaei N, Hawboldt K (2020) A critical review on life cycle analysis of algae biodiesel: current challenges and future prospects. Renew Sustain Energy Rev 134:110143. https://doi.org/10.1016/j.rser.2020.110143
Chau CK, Leung TM, Ng WY (2015) A review on Life Cycle Assessment, Life Cycle Energy Assessment and Life Cycle Carbon Emissions Assessment on buildings. Appl Energy 143:395–413. https://doi.org/10.1016/j.apenergy.2015.01.023
Chen W-H, Wang C-W, Ong HC et al (2019) Torrefaction, pyrolysis and two-stage thermodegradation of hemicellulose, cellulose and lignin. Fuel 258:116168. https://doi.org/10.1016/j.fuel.2019.116168
Cherubini F, Ulgiati S (2010) Crop residues as raw materials for biorefinery systems – A LCA case study. Appl Energy 87:47–57. https://doi.org/10.1016/j.apenergy.2009.08.024
Chopra J, Tiwari BR, Dubey BK, Sen R (2020) Environmental impact analysis of oleaginous yeast based biodiesel and biocrude production by life cycle assessment. J Clean Prod https://doi.org/10.1016/j.jclepro.2020.122349
Chung ZL, Tan YH, Chan YS et al (2019) Life cycle assessment of waste cooking oil for biodiesel production using waste chicken eggshell derived CaO as catalyst via transesterification. Biocatal Agric Biotechnol 21:101317. https://doi.org/10.1016/j.bcab.2019.101317
Collotta M, Champagne P, Tomasoni G et al (2019) Critical indicators of sustainability for biofuels: an analysis through a life cycle sustainabilty assessment perspective. Renew Sustain Energy Rev 115:109358. https://doi.org/10.1016/j.rser.2019.109358
Costa D, Jesus J, Virgínio e Silva J, Silveira M (2018) Life Cycle Assessment of Bioethanol Production from Sweet Potato (Ipomoea batatas L.) in an experimental plant. Bioenergy Res 11:715–725. https://doi.org/10.1007/s12155-018-9932-1
Deshmukh S, Kumar R, Bala K (2019) Microalgae biodiesel: a review on oil extraction, fatty acid composition, properties and effect on engine performance and emissions. Fuel Process Technol 191:232–247. https://doi.org/10.1016/j.fuproc.2019.03.013
Dominguez-Faus R, Powers SE, Burken JG, Alvarez PJ (2009) The Water Footprint of Biofuels: a drink or drive issue? Environ Sci Technol 43:3005–3010. https://doi.org/10.1021/es802162x
EASAC (2012) The current status of biofuels in the European Union, their environmental impacts and future prospects
Eggemann L, Escobar N, Peters R et al (2020) Life cycle assessment of a small-scale methanol production system: a power-to-fuel strategy for biogas plants. J Clean Prod 271:122476. https://doi.org/10.1016/j.jclepro.2020.122476
Ekener E, Hansson J, Gustavsson M (2018) Addressing positive impacts in social LCA—discussing current and new approaches exemplified by the case of vehicle fuels. Int J Life Cycle Assess 23:556–568. https://doi.org/10.1007/s11367-016-1058-0
Ekener-Petersen E, Höglund J, Finnveden G (2014) Screening potential social impacts of fossil fuels and biofuels for vehicles. Energy Policy 73:416–426. https://doi.org/10.1016/j.enpol.2014.05.034
Fargione J, Hill J, Tilman D et al (2008) Land Clearing and the Biofuel Carbon Debt. Sci (1979) 319:1235–1238. https://doi.org/10.1126/science.1152747

Fawzy S, Osman AI, Doran J, Rooney DW (2020) Strategies for mitigation of climate change: a review. Environ Chem Lett 18:2069–2094. https://doi.org/10.1007/s10311-020-01059-w

Fernández-Tirado F, Parra-López C, Romero-Gámez M (2016) Life cycle assessment of biodiesel in Spain: comparing the environmental sustainability of spanish production versus argentinean imports. Energy Sustain Dev 33:36–52. https://doi.org/10.1016/j.esd.2016.04.005

Firoz S (2008) A review: advantages and disadvantages of biodiesel. Int Res J Eng Technol 4(11):530–533

Foteinis S, Chatzisymeon E, Litinas A, Tsoutsos T (2020) Used-cooking-oil biodiesel: life cycle assessment and comparison with first- and third-generation biofuel. Renew Energy 153:588–600. https://doi.org/10.1016/j.renene.2020.02.022

Gallejones P, Pardo G, Aizpurua A, del Prado A (2015) Life cycle assessment of first-generation biofuels using a nitrogen crop model. Sci Total Environ 505:1191–1201. https://doi.org/10.1016/j.scitotenv.2014.10.061

Gandam PK, Chinta ML, Pabbathi NPP et al (2022) Second-generation bioethanol production from corncob – a comprehensive review on pretreatment and bioconversion strategies, including techno-economic and lifecycle perspective. Ind Crops Prod 186:115245. https://doi.org/10.1016/j.indcrop.2022.115245

Gerbens-Leenes W, Hoekstra AY (2010) Burning water: The water footprint of biofuel-based transport.

Gerbens-Leenes W, Hoekstra AY, van der Meer TH (2009) The water footprint of bioenergy. Proc Natl Acad Sci 106:10219–10223. https://doi.org/10.1073/pnas.0812619106

Gnansounou E, Kenthorai Raman J (2016) Life cycle assessment of algae biodiesel and its co-products. Appl Energy 161:300–308. https://doi.org/10.1016/j.apenergy.2015.10.043

Gnansounou E, Vaskan P, Pachón ER (2015) Comparative techno-economic assessment and LCA of selected integrated sugarcane-based biorefineries. Bioresour Technol 196:364–375. https://doi.org/10.1016/j.biortech.2015.07.072

González-García S, Gasol CM, Gabarrell X et al (2010a) Environmental profile of ethanol from poplar biomass as transport fuel in Southern Europe. Renew Energy 35:1014–1023. https://doi.org/10.1016/j.renene.2009.10.029

González-García S, Moreira MT, Feijoo G (2010b) Environmental performance of lignocellulosic bioethanol production from Alfalfa stems. Biofuels Bioprod Biorefin 4:118–131. https://doi.org/10.1002/bbb.204

González-García S, García-Rey D, Hospido A (2013a) Environmental life cycle assessment for rapeseed-derived biodiesel. Int J Life Cycle Assess 18:61–76. https://doi.org/10.1007/s11367-012-0444-5

González-García S, García-Rey D, Hospido A (2013b) Environmental life cycle assessment for rapeseed-derived biodiesel. Int J Life Cycle Assess 18:61–76. https://doi.org/10.1007/s11367-012-0444-5

Guinée JB, Heijungs R, Vijver MG, Peijnenburg WJGM (2017) Setting the stage for debating the roles of risk assessment and life-cycle assessment of engineered nanomaterials. Nat Nanotechnol 12:727–733. https://doi.org/10.1038/nnano.2017.135

Gunarathne V, Ashiq A, Ramanayaka S et al (2019) Biochar from municipal solid waste for resource recovery and pollution remediation. Environ Chem Lett 17:1225–1235. https://doi.org/10.1007/s10311-019-00866-0

Gupta R, McRoberts R, Yu Z et al (2022) Life cycle assessment of biodiesel production from rapeseed oil: influence of process parameters and scale. Bioresour Technol 360:127532. https://doi.org/10.1016/j.biortech.2022.127532

Hammond GP, Li B (2016) Environmental and resource burdens associated with world biofuel production out to 2050: footprint components from carbon emissions and land use to waste arisings and water consumption. GCB Bioenergy 8:894–908. https://doi.org/10.1111/gcbb.12300

Harris ZM, Spake R, Taylor G (2015) Land use change to bioenergy: a meta-analysis of soil carbon and GHG emissions. Biomass Bioenergy 82:27–39. https://doi.org/10.1016/j.biombioe.2015.05.008

Harris TM, Hottle TA, Soratana K et al (2016) Life cycle assessment of sunflower cultivation on abandoned mine land for biodiesel production. J Clean Prod 112:182–195. https://doi.org/10.1016/j.jclepro.2015.09.057

Hassan NS, Jalil AA, Hitam CNC et al (2020) Biofuels and renewable chemicals production by catalytic pyrolysis of cellulose: a review. Environ Chem Lett 18:1625–1648. https://doi.org/10.1007/s10311-020-01040-7

Hosseinzadeh-Bandbafha H, Nazemi F, Khounani Z et al (2022) Safflower-based biorefinery producing a broad spectrum of biofuels and biochemicals: a life cycle assessment perspective. Sci Total Environ 802:149842. https://doi.org/10.1016/j.scitotenv.2021.149842

Humpenöder F, Schaldach R, Cikovani Y, Schebek L (2013) Effects of land-use change on the carbon balance of 1st generation biofuels: an analysis for the European Union combining spatial modeling and LCA. Biomass Bioenergy 56:166–178. https://doi.org/10.1016/j.biombioe.2013.05.003

IEA (2021) World Energy Balances - Data product - IEA. https://www.iea.org/data-and-statistics/data-product/world-energy-balances. Accessed 1 Oct 2022

ILCD (2010) Analysis of existing Environmental Impact Assessment methodologies for use in Life Cycle Assessment (LCA). In: European Commission. https://eplca.jrc.ec.europa.eu/ilcd.html. Accessed 5 Oct 2022

IPCC (2022) Sixth Assessment Report: Mitigation of Climate Change, Chap. 2: Emissions trends and drivers. Geneva, Switzerland

Isler-Kaya A, Karaosmanoglu F (2022) Life cycle assessment of safflower and sugar beet molasses-based biofuels. Renew Energy 201:1127–1138. https://doi.org/10.1016/j.renene.2022.11.041

ISO (2006b) ISO - ISO 14044:2006 - Environmental management — Life cycle assessment — Requirements and guidelines. In: International Organization for Standardization (ISO). https://www.iso.org/standard/38498.html. Accessed 5 Oct 2022

ISO (2006a) ISO - ISO 14040:2006 - Environmental management — Life cycle assessment — Principles and framework. In: International Organization for Standardization (ISO). https://www.iso.org/standard/37456.html. Accessed 5 Oct 2022

Jain M, Chandrakant U, Orsat V, Raghavan V (2018) A review on assessment of biodiesel production methodologies from Calophyllum inophyllum seed oil. Ind Crops Prod 114:28–44. https://doi.org/10.1016/j.indcrop.2018.01.051

Jeswani HK, Falano T, Azapagic A (2015) Life cycle environmental sustainability of lignocellulosic ethanol produced in integrated thermo-chemical biorefineries. Biofuels Bioprod Biorefin 9:661–676. https://doi.org/10.1002/bbb.1558

Jeswani HK, Chilvers A, Azapagic A (2020) Environmental sustainability of biofuels: a review: Environmental sustainability of biofuels. Proc Royal Soc a: Math, Phys Eng Sci 476(2243):20200351

Jiang P, Zhao G, Liu L et al (2022) A negative-carbon footprint process with mixed biomass feedstock maximizes conversion efficiency, product value and CO2 mitigation. Bioresour Technol. https://doi.org/10.1016/j.biortech.2022.127004

Jolliet O, Margni M, Charles R et al (2003) IMPACT 2002+: a new life cycle impact assessment methodology. Int J Life Cycle Assess 8:324–330. https://doi.org/10.1007/BF02978505

Kalaivani K, Ravikumar G, Balasubramanian N (2014) Environmental impact studies of biodiesel production from Jatropha curcas

in India by life cycle assessment. Environ Prog Sustain Energy 33:1340–1349. https://doi.org/10.1002/ep.11913
Kaloudas D, Pavlova N, Penchovsky R (2021) Lignocellulose, algal biomass, biofuels and biohydrogen: a review. Environ Chem Lett 19:2809–2824. https://doi.org/10.1007/s10311-021-01213-y
Kargbo H, Harris JS, Phan AN (2021) Drop-in" fuel production from biomass: critical review on techno-economic feasibility and sustainability. Renew Sustain Energy Rev 135:110168. https://doi.org/10.1016/j.rser.2020.110168
Karlsson H, Börjesson P, Hansson P-A, Ahlgren S (2014) Ethanol production in biorefineries using lignocellulosic feedstock – GHG performance, energy balance and implications of life cycle calculation methodology. J Clean Prod 83:420–427. https://doi.org/10.1016/j.jclepro.2014.07.029
Khanna M, Crago CL, Black M (2011) Can biofuels be a solution to climate change? The implications of land use change-related emissions for policy. Interface Focus 1:233–247. https://doi.org/10.1098/rsfs.2010.0016
Khoo CG, Dasan YK, Lam MK, Lee KT (2019) Algae biorefinery: review on a broad spectrum of downstream processes and products. Bioresour Technol 292:121964
Kim S, Dale BE (2008) Life cycle assessment of fuel ethanol derived from corn grain via dry milling. Bioresour Technol 99:5250–5260. https://doi.org/10.1016/j.biortech.2007.09.034
Kovacs E, Hoaghia M-A, Senila L et al (2022). Life Cycle Assessment of Biofuels production processes in viticulture in the Context of Circular Economy. https://doi.org/10.3390/agronomy12061320
Kumar S, Singh J, Nanoti SM, Garg MO (2012) A comprehensive life cycle assessment (LCA) of Jatropha biodiesel production in India. Bioresour Technol 110:723–729. https://doi.org/10.1016/j.biortech.2012.01.142
Kumar Sharma A, Kumar Ghodke P, Manna S, Chen W-H (2021) Emerging technologies for sustainable production of biohydrogen production from microalgae: a state-of-the-art review of upstream and downstream processes. Bioresour Technol 342:126057. https://doi.org/10.1016/j.biortech.2021.126057
Lamb WF, Wiedmann T, Pongratz J et al (2021) A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. Environ Res Lett 16:073005
Lan K, Ou L, Park S et al (2020) Life cycle analysis of decentralized Preprocessing Systems for fast pyrolysis biorefineries with blended feedstocks in the Southeastern United States. Energy Technol 8:1900850. https://doi.org/10.1002/ente.201900850
Lapola DM, Schaldach R, Alcamo J et al (2010) Indirect land-use changes can overcome carbon savings from biofuels in Brazil. Proc Nat Acad Sci 107:3388–3393. https://doi.org/10.1073/pnas.0907318107
Lee KT, Ofori-Boateng C (2013) Life Cycle Assessment of Biodiesel from Palm Oil. In: Singh A, Pant D, Olsen SI (eds) Life Cycle Assessment of renewable energy sources. Springer London, London, pp 95–129
Li G, Lu Z, Zhang J et al (2020) Life cycle assessment of biofuel production from microalgae cultivated in anaerobic digested wastewater. Int J Agricultural Biol Eng 13:241–246. https://doi.org/10.25165/J.IJABE.20201301.4178
Lim S, Lee KT (2011) Parallel production of biodiesel and bioethanol in palm-oil-based biorefineries: life cycle assessment on the energy and greenhouse gases emissions. Biofuels Bioprod Biorefin 5:132–150. https://doi.org/10.1002/bbb.271
Mandade P, Bakshi BR, Yadav GD (2015) Ethanol from indian agro-industrial lignocellulosic biomass—a life cycle evaluation of energy, greenhouse gases, land and water. Int J Life Cycle Assess 20:1649–1658. https://doi.org/10.1007/s11367-015-0966-8
Martínez-Blanco J, Antón A, Rieradevall J et al (2011) Comparing nutritional value and yield as functional units in the environmental assessment of horticultural production with organic or mineral fertilization. Int J Life Cycle Assess 16:12–26. https://doi.org/10.1007/s11367-010-0238-6
Martinez-Hernandez E, Cui X, Scown CD et al (2019) Techno-economic and greenhouse gas analyses of lignin valorization to eugenol and phenolic products in integrated ethanol biorefineries. Biofuels Bioprod Biorefin 13:978–993. https://doi.org/10.1002/bbb.1989
Mathioudakis V, Gerbens-Leenes PW, Van der Meer TH, Hoekstra AY (2017) The water footprint of second-generation bioenergy: a comparison of biomass feedstocks and conversion techniques. J Clean Prod 148:571–582. https://doi.org/10.1016/j.jclepro.2017.02.032
Mekonnen MM, Hoekstra AY (2011) The green, blue and grey water footprint of crops and derived crop products. Hydrol Earth Syst Sci 15:1577–1600. https://doi.org/10.5194/hess-15-1577-2011
Michiels F, Geeraerd A (2020) How to decide and visualize whether uncertainty or variability is dominating in life cycle assessment results: a systematic review. Environ Model Softw 133:104841. https://doi.org/10.1016/j.envsoft.2020.104841
Mohd YUSOFSJH, Roslan AM, Ibrahim KN et al (2019) Life Cycle Assessment for Bioethanol Production from Oil Palm Frond Juice in an oil palm based Biorefinery. Sustainability 11:6928. https://doi.org/10.3390/su11246928
Mu D, Min M, Krohn B et al (2014) Life Cycle Environmental Impacts of Wastewater-Based Algal Biofuels. Environ Sci Technol 48:11696–11704. https://doi.org/10.1021/es5027689
Mukhopadhyay P, Chakraborty R (2021) LCA of sustainable biodiesel production from fried Borassus flabellifer oil in energy-proficient reactors: impact assessment of multi fuel-additives on pour point. Sustain Energy Technol Assess, NOx and engine performance. https://doi.org/10.1016/j.seta.2021.100994
Munagala M, Shastri Y, Nagarajan S, Ranade V (2022) Production of Bio-CNG from sugarcane bagasse: commercialization potential assessment in indian context. Ind Crops Prod 188:115590. https://doi.org/10.1016/j.indcrop.2022.115590
Muñoz I, Flury K, Jungbluth N et al (2014) Life cycle assessment of bio-based ethanol produced from different agricultural feedstocks. Int J Life Cycle Assess 19:109–119. https://doi.org/10.1007/s11367-013-0613-1
Ocreto JB, Chen W-H, Ubando AT et al (2021) A critical review on second- and third-generation bioethanol production using microwaved-assisted heating (MAH) pretreatment. Renew Sustain Energy Rev 152:111679. https://doi.org/10.1016/j.rser.2021.111679
Olukoya IA, Ramachandriya KD, Wilkins MR, Aichele CP (2014) Life cycle assessment of the production of ethanol from eastern redcedar. Bioresour Technol 173:239–244. https://doi.org/10.1016/j.biortech.2014.09.100
Osman AI, Hefny M, Abdel Maksoud MIA et al (2021a) Recent advances in carbon capture storage and utilisation technologies: a review. Environ Chem Lett 19:797–849. https://doi.org/10.1007/s10311-020-01133-3
Osman AI, Mehta N, Elgarahy AM et al (2021b) Conversion of biomass to biofuels and life cycle assessment: a review. Environ Chem Lett 19:4075–4118. https://doi.org/10.1007/s10311-021-01273-0
Ou X, Zhang X, Chang S, Guo Q (2009) Energy consumption and GHG emissions of six biofuel pathways by LCA in (the) people's Republic of China. Appl Energy 86:S197–S208. https://doi.org/10.1016/j.apenergy.2009.04.045
Owsianiak M, Laurent A, Bjørn A, Hauschild MZ (2014) IMPACT 2002+, ReCiPe 2008 and ILCD's recommended practice for characterization modelling in life cycle impact assessment: a case study-based comparison. Int J Life Cycle Assess 19:1007–1021. https://doi.org/10.1007/s11367-014-0708-3

Passell H, Dhaliwal H, Reno M et al (2013) Algae biodiesel life cycle assessment using current commercial data. J Environ Manage 129:103–111. https://doi.org/10.1016/j.jenvman.2013.06.055

Patel K, Singh SK (2022) Assessing the sustainability of municipal solid waste management using life-cycle analysis. Proceedings of the Institution of Civil Engineers - Waste and Resource Management 175:64–74. https://doi.org/10.1680/jwarm.21.00018

Patel K, Singh SK (2022b) A life cycle approach to environmental assessment of wastewater and sludge treatment processes. Water and Environment Journal 36:412–424. https://doi.org/10.1111/wej.12774

Patel K, Singh SK (2023) In: Al Khaddar R, Singh SK, Kaushika ND et al (eds) Assessing the sustainability of Jatropha and Rapeseed Biodiesel: An LCA Approach BT - recent developments in Energy and Environmental Engineering. Springer Nature Singapore, Singapore, pp 191–200

Peng L, Fu D, Chu H et al (2020) Biofuel production from microalgae: a review. Environ Chem Lett 18:285–297. https://doi.org/10.1007/s10311-019-00939-0

Perianes-Rodriguez A, Waltman L, van Eck NJ (2016) Constructing bibliometric networks: a comparison between full and fractional counting. J Informetr 10:1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

Pleanjai S, Gheewala SH (2009) Full chain energy analysis of biodiesel production from palm oil in Thailand. Appl Energy 86:S209–S214. https://doi.org/10.1016/j.apenergy.2009.05.013

Pradhan P, Karan P, Chakraborty R (2022) Life cycle sustainability assessment of optimized biodiesel production from used rice bran oil employing waste derived-hydroxyapatite supported vanadium catalyst. Environ Sci Pollut Res 29:20064–20077. https://doi.org/10.1007/s11356-021-16482-x

Pragya N, Pandey KK (2016) Life cycle assessment of green diesel production from microalgae. Renew Energy 86:623–632. https://doi.org/10.1016/j.renene.2015.08.064

RAE (2017) Sustainability of liquid biofuels. Royal Academy of Engineering

Reaño RL (2020) Assessment of environmental impact and energy performance of rice husk utilization in various biohydrogen production pathways. Bioresour Technol 299:122590. https://doi.org/10.1016/j.biortech.2019.122590

Rocha MH, Capaz RS, Lora EES et al (2014) Life cycle assessment (LCA) for biofuels in brazilian conditions: a meta-analysis. Renew Sustain Energy Rev 37:435–459. https://doi.org/10.1016/j.rser.2014.05.036

RSB, (2020) The Viability of South African Sugarcane Ethanol as Feedstock for Sustainable aviation fuel production Part III GHG Emissions

Saranya G, Ramachandra TV (2020) Life cycle assessment of biodiesel from estuarine microalgae. Energy Convers Management: X 8:100065. https://doi.org/10.1016/j.ecmx.2020.100065

Sawatdeenarunat C, Surendra KC, Takara D et al (2015) Anaerobic digestion of lignocellulosic biomass: Challenges and opportunities. Bioresour Technol 178:178–186

Schmidt Rivera XC, Gallego-Schmid A, Najdanovic-Visak V, Azapagic A (2020) Life cycle environmental sustainability of valorisation routes for spent coffee grounds: from waste to resources. Resour Conserv Recycl 157:104751. https://doi.org/10.1016/j.resconrec.2020.104751

Sebastião D, Gonçalves MS, Marques S et al (2016) Life cycle assessment of advanced bioethanol production from pulp and paper sludge. Bioresour Technol 208:100–109. https://doi.org/10.1016/j.biortech.2016.02.049

Sharma T, Dasgupta D, Singh J et al (2020) Yeast lipid-based biofuels and oleochemicals from lignocellulosic biomass: life cycle impact assessment. Sustain Energy Fuels 4:387–398. https://doi.org/10.1039/C9SE00540D

Shirvani T, Yan X, Inderwildi OR et al (2011) Life cycle energy and greenhouse gas analysis for algae-derived biodiesel. Energy Environ Sci 4:3773–3778. https://doi.org/10.1039/C1EE01791H

Silalertruksa T, Pongpat P, Gheewala SH (2017) Life cycle assessment for enhancing environmental sustainability of sugarcane biorefinery in Thailand. J Clean Prod 140:906–913. https://doi.org/10.1016/j.jclepro.2016.06.010

Silitonga AS, Masjuki HH, Ong HC et al (2016) Pilot-scale production and the physicochemical properties of palm and Calophyllum inophyllum biodiesels and their blends. J Clean Prod 126:654–666. https://doi.org/10.1016/j.jclepro.2016.03.057

Sims R, Schaeffer R (2014) Chap. 8 Transport. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change

Sinha S, Suzuki S, Kojima T et al (2013) Life Cycle Analysis and Modelling (LCAM) of Jatropha as Biofuel in Dynamic Economic Environment of newly emerging economies, Center for Asia-Pacific Studies. Rev Asian Pacific Stud 38:119–139

Siqueira SF, Deprá C, Zepka LQ, Jacob-Lopes E (2018) Life Cycle Assessment (LCA) of third-generation Biodiesel Produced Heterotrophically by Phormidium Autumnale. Open Biotechnol J 12:270–281. https://doi.org/10.2174/1874070701812010270

Soam S, Kumar R, Gupta RP et al (2015) Life cycle assessment of fuel ethanol from sugarcane molasses in northern and western India and its impact on indian biofuel programme. Energy 83:307–315. https://doi.org/10.1016/j.energy.2015.02.025

Soam S, Kapoor M, Kumar R et al (2016) Global warming potential and energy analysis of second generation ethanol production from rice straw in India. Appl Energy 184:353–364. https://doi.org/10.1016/j.apenergy.2016.10.034

Soratana K, Harper WF Jr, Landis AE (2012) Microalgal biodiesel and the renewable fuel Standard's greenhouse gas requirement. Energy Policy 46:498–510. https://doi.org/10.1016/j.enpol.2012.04.016

Soratana K, Barr WJ, Landis AE (2014) Effects of co-products on the life-cycle impacts of microalgal biodiesel. Bioresour Technol 159:157–166. https://doi.org/10.1016/j.biortech.2014.02.070

Sousa VMZ, Luz SM, Caldeira-Pires A et al (2017) Life cycle assessment of biodiesel production from beef tallow in Brazil. Int J Life Cycle Assess 22:1837–1850. https://doi.org/10.1007/s11367-017-1396-6

Spatari S, Larnaudie V, Mannoh I et al (2020) Environmental, exergetic and economic tradeoffs of catalytic- and fast pyrolysis-to-renewable diesel. Renew Energy 162:371–380. https://doi.org/10.1016/j.renene.2020.08.042

Spinelli D, Jez S, Pogni R, Basosi R (2013) Environmental and life cycle analysis of a biodiesel production line from sunflower in the Province of Siena (Italy). Energy Policy 59:492–506. https://doi.org/10.1016/j.enpol.2013.04.009

Sreekumar A, Shastri Y, Wadekar P et al (2020) Life cycle assessment of ethanol production in a rice-straw-based biorefinery in India. Clean Technol Environ Policy 22:409–422. https://doi.org/10.1007/s10098-019-01791-0

Stephenson AL, Dupree P, Scott SA, Dennis JS (2010) The environmental and economic sustainability of potential bioethanol from willow in the UK. Bioresour Technol 101:9612–9623. https://doi.org/10.1016/j.biortech.2010.07.104

Sun L-H, Wang Y-Y, Gong Y-Q (2022) Life cycle assessment of rice bran oil production: a case study in China. Environ Sci Pollut Res 29:39847–39859. https://doi.org/10.1007/s11356-021-18172-0

Ubando AT, Anderson S, Ng E, Chen W-H et al (2022) Life cycle assessment of microalgal biorefinery: a state-of-the-art review. Bioresour Technol 360:127615. https://doi.org/10.1016/j.biortech.2022.127615

Uppalapati S, Jani SP, Khan MBA et al (2022) A comparative assessment on life cycle analysis of the biodiesel fuels produced from

soybean, Jatropha, Calophyllum inophyllum, and microalgae. Energy Sources Part A: Recovery Utilization and Environmental Effects 44:3253–3272. https://doi.org/10.1080/15567036.2022.2053248

USDA (2021a) Biofuels Annual. Sao Paulo ATO. United State Department of Agriculture, Foreign Agricultural Service, Brazil

USDA (2021b) Biofuels Annual. New Delhi. United States Department of Agriculture, Foreign Agricultural Service, India

Uusitalo V, Väisänen S, Havukainen J et al (2014) Carbon footprint of renewable diesel from palm oil, jatropha oil and rapeseed oil. Renew Energy 69:103–113. https://doi.org/10.1016/j.renene.2014.03.020

van Dam J, Faaij APC, Hilbert J et al (2009) Large-scale bioenergy production from soybeans and switchgrass in Argentina: part A: potential and economic feasibility for national and international markets. Renew Sustain Energy Rev 13:1710–1733. https://doi.org/10.1016/j.rser.2009.03.009

Vrech A, Ferfuia C, Bessong Ojong W et al (2019) Energy and environmental sustainability of Jatropha-Biofuels Chain from nontoxic accessions in Cameroon. Environ Prog Sustain Energy 38:305–314. https://doi.org/10.1002/ep.12928

Wang E, Shen Z (2013) A hybrid Data Quality Indicator and statistical method for improving uncertainty analysis in LCA of complex system – application to the whole-building embodied energy analysis. J Clean Prod 43:166–173. https://doi.org/10.1016/j.jclepro.2012.12.010

Wang L, Littlewood J, Murphy RJ (2013a) Environmental sustainability of bioethanol production from wheat straw in the UK. Renew Sustain Energy Rev 28:715–725. https://doi.org/10.1016/j.rser.2013.08.031

Wang M, Shi Y, Xia X et al (2013b) Life-cycle energy efficiency and environmental impacts of bioethanol production from sweet potato. Bioresour Technol 133:285–292. https://doi.org/10.1016/j.biortech.2013.01.067

Wu W, Lei Y-C, Chang J-S (2019) Life cycle assessment of upgraded microalgae-to-biofuel chains. Bioresour Technol 288:121492. https://doi.org/10.1016/j.biortech.2019.121492

Xie X, Zhang T, Wang L, Huang Z (2017) Regional water footprints of potential biofuel production in China. Biotechnol Biofuels 10:95. https://doi.org/10.1186/s13068-017-0778-0

Yan X, Crookes RJ (2009) Life cycle analysis of energy use and greenhouse gas emissions for road transportation fuels in China. Renew Sustain Energy Rev 13:2505–2514. https://doi.org/10.1016/j.rser.2009.06.012

Yang Y, Zhang B, Cheng J, Pu S (2015) Socio-economic impacts of algae-derived biodiesel industrial development in China: an input–output analysis. Algal Res 9:74–81. https://doi.org/10.1016/j.algal.2015.02.010

Yang Q, Mašek O, Zhao L et al (2021) Country-level potential of carbon sequestration and environmental benefits by utilizing crop residues for biochar implementation. Appl Energy 282:116275. https://doi.org/10.1016/j.apenergy.2020.116275

Yazan DM, Mandras G, Garau G (2017) Environmental and economic sustainability of integrated production in bio-refineries: the thistle case in Sardinia. Renew Energy 102:349–360. https://doi.org/10.1016/j.renene.2016.10.055

Zhang J, Jia C, Wu Y et al (2017) Life cycle energy efficiency and environmental impact assessment of bioethanol production from sweet potato based on different production modes. PLoS ONE 12(7): e0180685

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Evaluating Metaheuristic Algorithms and Neural Networks for PID Tuning

Abhigyan Adarsh
Dept. of Electrical Engineering
Delhi Technological University
Delhi, India
abhigyanadarsh_2k19ee005@dtu.ac.in

Amal Chawla
Dept. of Electrical Engineering
Delhi Technological University
Delhi, India
amalchawla_2k19ee032@dtu.ac.in

Anirudh Singh
Dept. of Electrical Engineering
Delhi Technological University
Delhi, India
anirudhsingh_2k19ee042@dtu.ac.in

Bharat Bhushan
Dept of Electrical Engineering
Delhi Technological University
Delhi, India
bharat@dce.ac.in

2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES) | 979-8-3503-4729-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSSES58299.2023.10200433

***Abstract*—This paper aims at comparing various PID tuning methodologies. An armature-controlled DC motor has been modeled in Simulink using its actual parameters for the study. The metaheuristic approaches i.e., Flower Pollination Algorithm and Grey Wolf Optimization have been coded on MATLAB and the tuned parameters have been imported on Simulink to check their response of the system. These algorithms work on ITAE as their objective function. ANFIS controller for tuning PID values is implemented on Simulink. Its training data is imported from Fuzzy Logic Controller's tuned parameters. Finally, the results of the controllers are analyzed and distinguished by values of Rise Time (Tr), Peak Overshoot (Mp), and Peak Time (Tp).**

***Index Terms*—DC Motor, SIMULINK, PID, GWO, FPA, ITAE, Objective function, ANFIS, Fuzzy logic.**

I. Introduction

Engineers rely on motion-control devices to improve efficiency and production rates on automated factory floors, or at least maintain them. One of the earliest devices designed to transform electrical power into mechanical power was the DC motor. Due to their high levels of dependability, versatility, and affordability[1], DC motors are frequently used when speed of motor and position control are essential in industrial settings, robot manipulators, and appliances.

P, PI, and PID tuning are methods for optimizing feedback control systems. Proportional (P) control uses the error to determine the control action, while PI control adds an integral term to eliminate steady-state errors. PID control further adds a derivative term to dampen oscillations and improve stability. PID controllers offer faster response, reduced steady-state error, improved stability, and adaptability to system changes compared to P or PI controllers.

Motor control applications frequently use PID controllers because of their straightforward topologies and easily understandable control algorithms. Proportional-integral-derivative (PID) control offers a simple and incredibly useful solution to many concerns in the real world due to its three-term structure. Both transient and steady-state behaviors may be governed by all this. As a result, PID controllers are frequently used in traditional speed control loops[2]. The drawbacks of conventional PID controllers include unwanted overshoot, slow response, sensitivity to controller gains, and rapid changes in load torque[3]. The varied behavior of the system parameters under operating environment and the noise that is present in the circuit are the two main issues with motor control[4]. As a result, it is difficult to use standard techniques for motor control of sophisticated, nonlinear, and time-varying systems. To achieve an ideal state under field conditions, tuning the PID controller's parameters is therefore obviously necessary[5].

The metaheuristic algorithms [6] like Flower Pollination Algorithm and Grey Wolf Optimization use ITAE as an objective function for optimization. This objective function is minimized using a global minimization approach. These algorithms are inspired by nature and approach optimization of PID irrespective of plant model, thus providing robust results.

When using advanced speed control approaches, such as Adaptive Neuro-Fuzzy Interface System (ANFIS), desired speed tracking with sufficient precision in the presence of abrupt interruptions and parameter fluctuation can be accomplished. The ability to make complicated nonlinear control actions, also for indefinite nonlinear systems, is a feature of a nonlinear controller that is endowed with fuzzy control theory [5]. These control techniques bring the complexity of Machine Learning and Neural Networks into the hardware domain of PID. These techniques give better results in any scenario.

This paper is organized into seven segments: Section II goes on to discuss the mathematical model; Section III aims to discuss the metaheuristic approach to PID tuning and objective functions; Section IV explores Grey Wolf Optimization; Section V explains the Flower Pollination Algorithm; Section VI mentions simulation results and discussions; Section VII analyses conclusions; and Section VIII concludes. These sections are followed by an acknowledgment and a list of references.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:10:35 UTC from IEEE Xplore. Restrictions apply.

II. Mathematical Model

The voltage applied to the motor can be changed while the field remains constant in separately excited DC Motors.

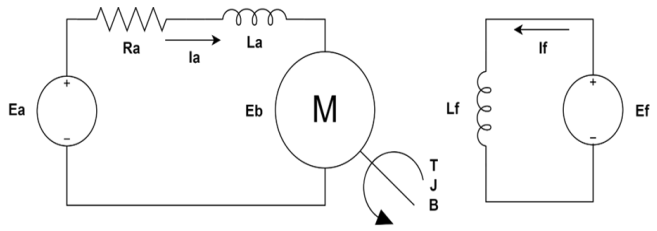


Figure 1. Equivalent Circuit of Armature Controlled DC Motor.

Table I
EQUIVALENT CIRCUIT PARAMETERS AND VALUES

Parameter	Nomenclature	Values
Moment of Inertia	J	$0.093Kg-m^2$
Friction Coefficient	B	$0.007N-ms$
Back Emf Constant	K_b	$0.6V/rad-s^-1$
Torque Constant	K_T	$0.7273Nm/A$
Armature Resistance	R_a	0.6
Armature Inductance	L_a	$0.006H$

The mathematical equations of the Armature Controlled DC Motor are as follows:

$$\Phi \propto i_f$$

$$\Phi = K_f i_f \tag{1}$$

Where,

$$T \propto \Phi i_a$$

$$T = K_a \Phi i_a = K_a K_f \Phi i_a$$

$$T = K_T i_a \tag{2}$$

$$E_b \propto \Phi \omega$$

$$E_b = K_b \omega \tag{3}$$

Applying KVL,

$$V = i_a R_a + L_a \frac{di_a}{dt} + E_b \tag{4}$$

And,

$$T = J\frac{d^2\theta}{dt^2} + B\frac{d\theta}{dt} \tag{5}$$

Load Torque,

$$T = J\frac{d^2\theta}{dt^2} + B\frac{d\theta}{dt} + T_L \tag{6}$$

By Laplace Transform,

$$V(s) = i_a(s)(R_a + sL_a) + E_b$$

$$V(s) - E_b(s) = i_a(s)(R_a + sL_a)$$

$$E_b(s) = K_b s\theta(s)$$

$$T(s) = Ki_a(s)$$

$$T(s) = \theta(s)(Js^2 + sB)$$

$$T(s) = \omega(s)(Js + B)$$

$$T(s) = K_T i_a I(s)$$

$$T(s) = \frac{\omega(s)}{V(s)} = \frac{K_T}{(R_a + sLa)(Js + B) + K_b K_T} \tag{7}$$

The transfer function of the DC Motor obtained after putting Equivalent circuit Parameters defined in Table I is as follows:

$$T(s) = \frac{1}{0.0007671s^2 + 0.076785s + 0.60654} \tag{8}$$

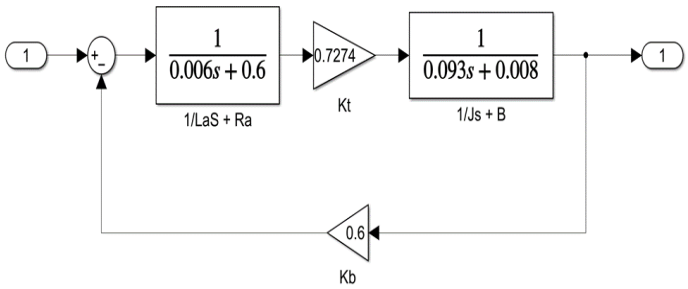


Figure 2. Block Diagram of DC Motor designed in SIMULINK

Figure 2, depicts the block diagram modeled in SIMULINK using Equation (7).

III. Metaheuristic approach of PID Tuning and Objective Functions

A. Metaheuristic approach of PID Tuning

Process control makes considerable use of metaheuristic algorithms [7] because of their quick response times, high optimization capabilities, and simple architectural design. Metaheuristic algorithms are more effective at resolving higher-dimensional optimization issues than conventional optimization techniques. They are easily adaptable to modern classical controller design approaches due to their durability. Metaheuristic algorithms can be used as a fundamental method for creating a uniform and enhanced set of resilient operational models that require structured controllers, regardless of the order in which the models are developed.

B. Objective Functions

The real-valued function whose value is to be either minimized or maximized over the range of feasible solutions is characterized as the objective function[8] in a mathematical optimization problem. In this case, our optimization goal function would be the Integral Time Absolute Error (ITAE)[9]: ITAE =$\int_0^\tau t|e(t)|$

IV. Grey Wolf Optimization (GWO)

A. Implementation

The grey wolf, which hunts massive prey in gangs and thrives on individual wolf coordination, inspired this algorithm. This behavior is notable for two reasons:

1) Social hierarchy
2) Hunting mechanism.

The grey wolf has a sophisticated social structure due to its gregarious nature. The phrase ”dominance hierarchies” refers

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:10:35 UTC from IEEE Xplore. Restrictions apply.

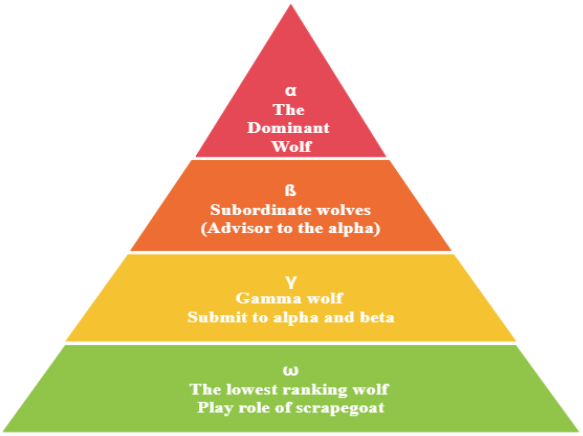


Figure 3. Grey Wolf Algorithm dominance hierarchies

to the hierarchical order depicted in Fig 3, in which wolves are ranked according to their authority and strength[10]. As a result, alphas, betas, deltas, and omegas exist.

1) At the top of the pack's hierarchy are the alpha male and female. Each member of the pack has been assigned a rank. The wolf's hierarchical order is about more than just dominance and aggression; it also assists vulnerable members of the pack who are unable to hunt for themselves.
2) The beta wolf comes next, who supports the alpha wolf's decisions and helps to preserve pack discipline.
3) The delta wolf is inferior to the beta wolf. They are frequently powerful, but lack the necessary leadership qualities and self-confidence to take on leadership responsibilities.
4) Eventually, the omega wolf wields little power, and other wolves will quickly chase him. Omega wolf is also in charge of monitoring the younger wolves.

At each level, we will use the previously described methods to indicate the three best solutions by alpha, beta, and delta, as well as alternative options by omega. This essentially means that the optimization process is guided by the flow of the three best solutions. In addition, the prey will be the best optimization solution[11].
Most of the logic follows the equations:

$$\overrightarrow{D} = \mid \overrightarrow{C}.\overrightarrow{X_p}(t) - \overrightarrow{X}(t) \mid \tag{9}$$

$$\overrightarrow{X}(t+1) = \mid \overrightarrow{X_p}(t) - \overrightarrow{A}.\overrightarrow{D} \mid \tag{10}$$

where t indicates the latest iteration, A and C are regression coefficients, Xp is the vector of the prey's movement, and X is the vector of the wolf's position. Cartesian coordinates A and B seem to be similar.

$$\overrightarrow{A} = 2\overrightarrow{a}.\overrightarrow{r_1} - \overrightarrow{a} \tag{11}$$

$$\overrightarrow{C} = 2\overrightarrow{r_2} \tag{12}$$

where a's components decline linearly from 2 to 0 over iterations and r1, r2 are random vectors with values ranging from [0, 1] computed for each wolf at each iteration[12]. Vector A governs the trade-off between exploration and exploitation, whereas vector C always includes some unpredictability. This is crucial because our agents can become stranded in local optima, which most metaheuristics can avoid. Because we don't know the true position of the optimal solution, Xp is determined by the three best answers.

B. PID Tuning using GWO

The tuning parameters of our PID controller, namely Kp, Ki, and Kd, will be modified by assessing the fitness of our DC motor system's objective function, namely ITAE (Integral Time Absolute Error). These parameters are then fed into our DC motor model's PID controller. Figure 5 shows a block schematic for a PID-fed DC motor with parameters optimised using the GWO Algorithm.

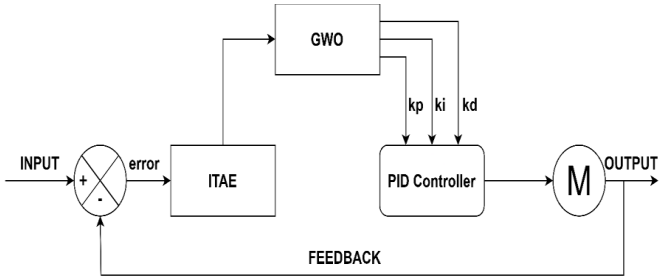


Figure 4. DC motor drive speed with PID tuning in closed loop control

The Grey Wolf Optimisation algorithm will be applied as defined in the flowchart shown in Fig 6 with the initialisation parameters being[13]:

1) Number of iterations.
2) Size of the population (search area).
3) Number of variables (Kp, Ki, Kd).

V. Flower Pollination Algorithm(FPA)

A. Algorithm

Pollen transmission is often a component of flower pollination[14], and it frequently involves pollinators which contain animals such as birds, and bugs, among others. In real world, a highly specialised flower-pollinator partnership has co-evolved as a result of specific blooms and insects. For optimal pollination, some blooms, for example, can only draw and depend on a certain species of insect. Pollinators, also referred to as pollen vectors, appear in great diversity. At least 200,000 different pollinators, including birds and insects[15], are thought to exist. A notable example of a pollinator is the honeybee, which has the capacity to achieve what is known as blossom constancy. In terms of behavior, these pollinators frequently visit only a few flower types while avoiding others. Floral constancy may have evolutionary benefits since it will improve the transmission of pollen to related or identical plants, maximizing the continuation of a particular genus of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:10:35 UTC from IEEE Xplore. Restrictions apply.

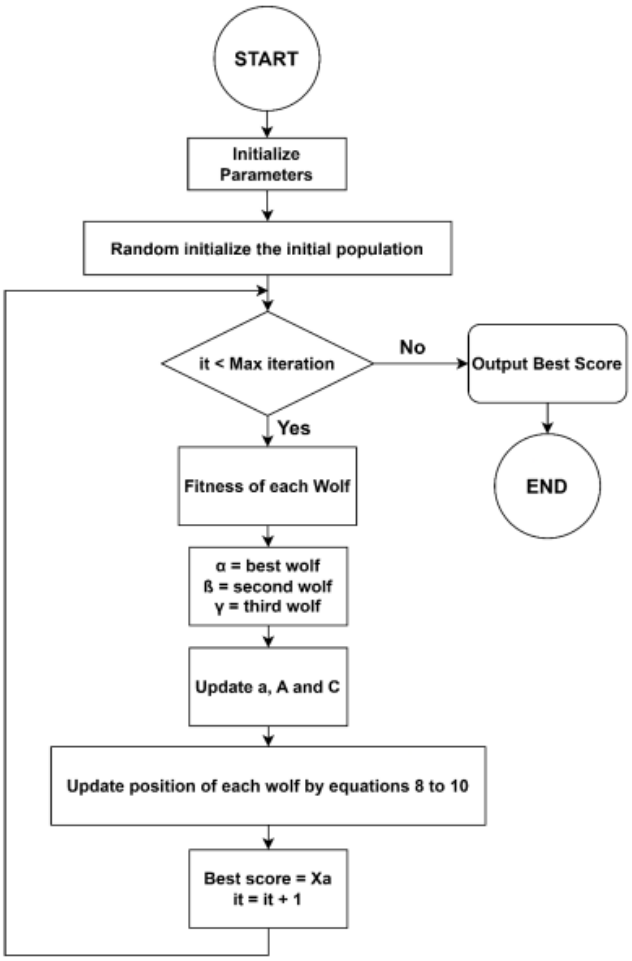


Figure 5. Flowchart of Grey Wolf Optimization Algorithm

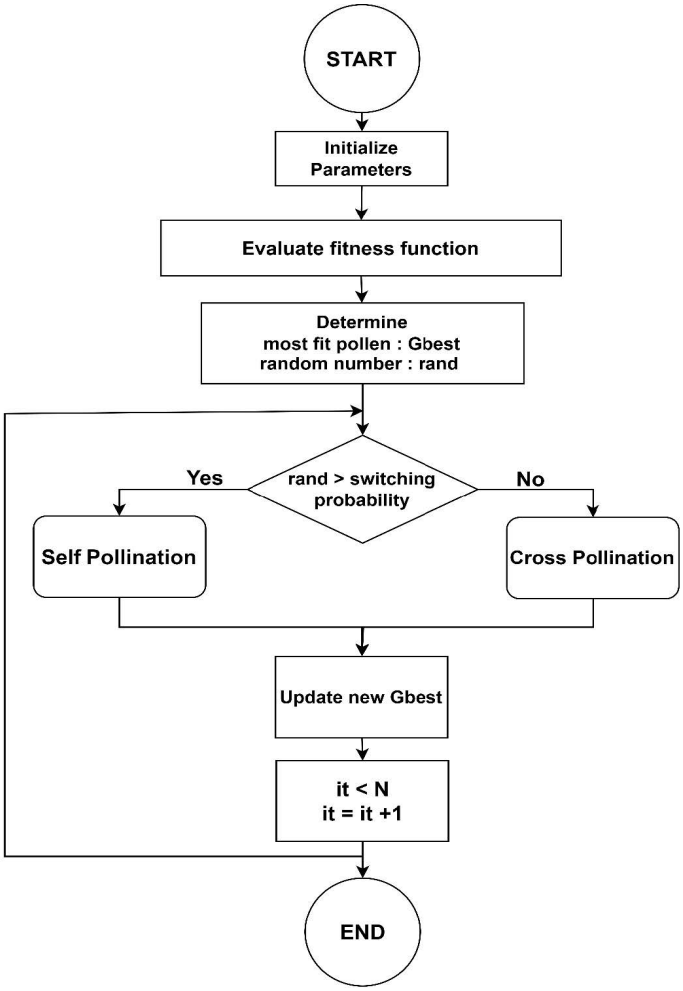


Figure 6. Flowchart of Flower Pollination Optimization Algorithm

flowers. Such floral stability can be beneficial. Figure 6 shows a block schematic for a PID-fed DC motor with parameters optimised using the Flower Pollination Algorithm[16,17].

B. PID Tuning using FPA

The values of the controller parameters for our analysis of the performance of DC motors[18] can be obtained by applying the flower pollination algorithm with K_p, K_i, K_d.

The following four guidelines are employed for simplicity[19]:

1) Lévy flights are followed by pollen-carrying pollinators, and biotic and cross-pollination are considered to be global pollination processes. (Rule 1).
2) Local pollination techniques include self-pollination and abiotic pollination. (Rule 2).
3) Pollinators like insects can develop flower constancy, which is compared to a reproduction likelihood proportionate to the similarity of the two flowers involved. (Rule 3).
4) Switch likelihood p[0,1] that is somewhat biased in favor of local pollination can be used to influence how local and global pollination interacts or switch. (Rule 4).

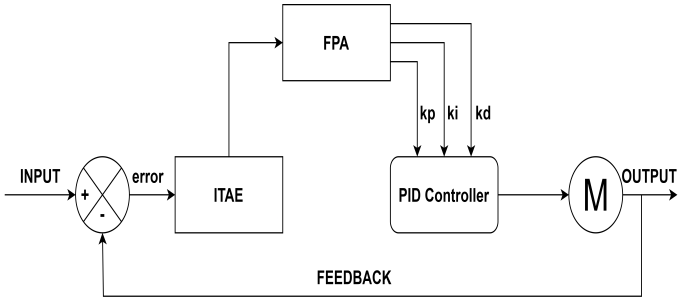


Figure 7. DC motor drive speed with PID tuning in closed loop control

VI. Adaptive Neuro Fuzzy Interface System (ANFIS)

A. ANFIS Controller

Fuzzy Logic and Neural Network together form ANFIS controller[20]. A number of inputs are applied to the neural network. The neural network produces several typical outputs based on its inputs. Neural Network is trained based on the input and output[21]. Fuzzy logic is used to produce the IF THEN rules and membership functions once the neural network has finished training

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:10:35 UTC from IEEE Xplore. Restrictions apply.

B. Data Sampling and Controller Training

In order to avoid making the validation process too simple, we must choose a data set that is both representational of the data the trained model is meant to imitate and sufficiently different from the training data set[22]. For the sampling process values of K_p are collected from the fuzzy controller. We have collected the data for 4 different signals (Figure 8). This data is used in the form of 2D array and data from these 4 signals are combined to form the training data for ANFIS controller for K_p. The same sampling process is used for the data collection of K_i and K_d. The sampled data collected is

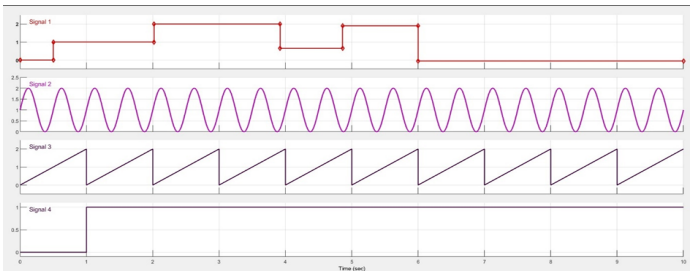


Figure 8. Signals used for training ANFIS Controller

now trained on the ANFIS controller. Pi type of membership function is used as it produced the best results in the case of Fuzzy tuning and 10 for each variable. The output type used is linear. The number of training iterations were 10.

C. Tuning of ANFIS Controller

After the training ANFIS controller for each of the K_p, K_i and K_d values were introduced in the model shown in Figure 9 as block diagram and the results were compared.

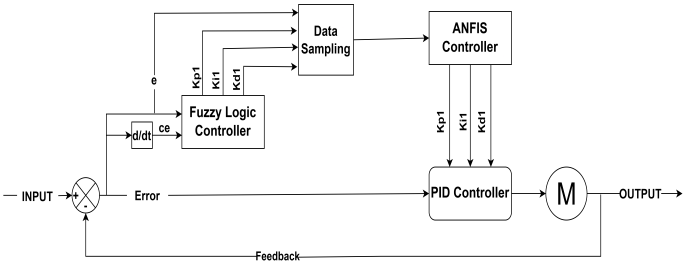


Figure 9. Block Diagram for tuning of ANFIS Controller

VII. Simulation Results and Discussion

1. Convergence of Objective Function

Figure 10 and Figure 11 demonstrates how objective function ITAE is converging for Grey Wolf Optimization and Flower Pollination ALgorithm run in MATLAB to obtain parameter for PID Controller.

2. Comparative Analysis of GWO, FPA, and ANFIS

Figure 12 shows the Combined Response of GWO, FPA, and ANFIS with respect to Step Input. Flower pollination can provide better results in every aspect as compared to Grey Wolf optimization. FPA tuning provide us with less overshoot and the system is becoming stable faster.

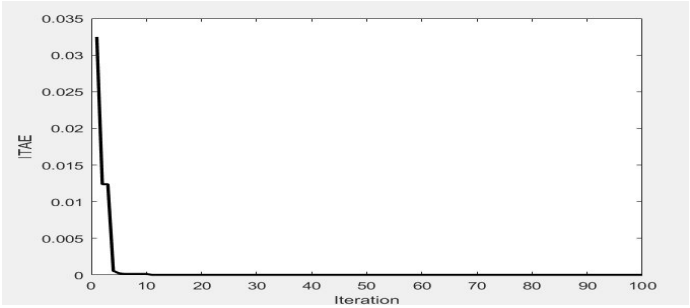


Figure 10. Convergence of Objective Function for GWO

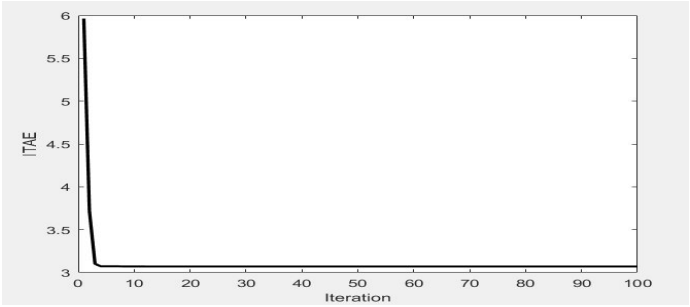


Figure 11. Convergence of Objective Function for FPA

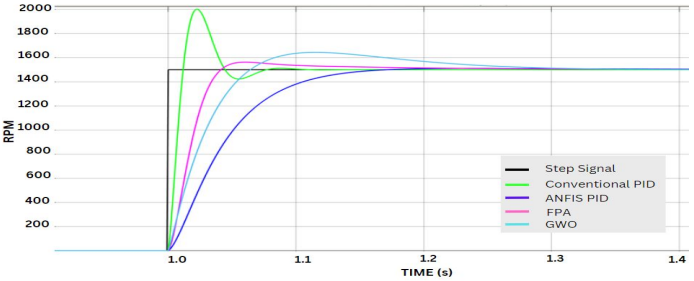


Figure 12. Comparative Analysis of all the Tuning Methodologies

Table II shows the results obtained by different Algorithms for various Parameters involving Peak Overshoot(M_p), Peak Time(T_p) and Rise Time(T_r).

Table II
System Characteristics for Tuned Parameters of FPA and GWO

Tuning Algorithm	Tr(ms)	Mp(%)	Tp(ms)
Conventional PID Tuning	10	33.4	22
Grey Wolf Optimisation	49	9.6	115
Flower Pollination	14	2.7	27

ANFIS controller has significant improvement in terms of peak overshoot. With the training done on 7500 data points computed from Fuzzy Logic Controller which in turn provides very optimized results, this was inevitable. The overall stabilization is also better. The ANFIS controller only lags in terms of rise time with respect to our metaheuristic algorithms. Due to a large amount of training and computation required.

Table 3 displays the parameters of the ANFIS controller with the pi membership function, and it demonstrates that ANFIS is better than Fuzzy for overshoot control, but it is slower since the rise time and peak time for ANFIS are considerably longer than for Fuzzy.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:10:35 UTC from IEEE Xplore. Restrictions apply.

Table III
SYSTEM CHARACTERISTICS FOR TUNED PARAMETERS OF ANFIS CONTROLLER

Membership Function	Tr(ms)	Mp(%)	Tp(ms)
Pi Shaped	104.8	0.39	555

VIII. CONCLUSION

Using MATLAB/Simulink, a DC motor is simulated and its speed is controlled. We have employed metaheuristic methods of optimization: Grey Wolf Optimisation, and Flower Pollination Algorithm. The nature-inspired metaheuristic approaches for tuning PID provide robust results and are irrespective of the plant model. The use of the objective function for optimization in these techniques provides better stability with changes. Thus, we have a minimum rise time for a plant model. Considering the unique algorithms we see faster settling in FPA, along with better rise time and also lesser overshoot. Hence, Flower pollination with cross-pollination and self-pollination provides a better methodology for tuning compared to conventional techniques. ANFIS being a neural network optimization technique is computation heavy. Thus the rise time in its case is more. Yet the settling time describes the efficiency it provides.

Hence both approaches to tuning have individual benefits. Metaheuristic algorithms are resistant to change (smaller rise time) while Neural networks are better in stabilization (smaller settling time and peak overshoot).

ACKNOWLEDGMENT

We would like to convey our sincere gratitude to Professor Bharat Bhushan of the Electrical Engineering Department, Delhi Technological University for providing us with the chance to conduct this research and for his helpful advice. His enthusiasm, vision, sincerity, and drive have greatly motivated us. It was a great honor and joy to be able to work and study under his leadership. We are immensely grateful for everything he has given us. He has always been an inspiration to me. We also like to express our gratitude to all of the teaching and support personnel at Delhi Technological University's Electrical Engineering Department for their direct and/or indirect contributions to the project.

REFERENCES

[1] T. Thu Lynn and E. Eaint, "Position Control of DC Servo Drive by Fuzzy Logic Controller in Flat-Bed Screen Printing Machine", ASRJETS-Journal, vol. 26, no. 1, pp. 8–19, Oct. 2016.

[2] A. Kiam Heong, G. Chong, and L. Yun, "PID control system analysis, design, and technology," IEEE Transactions on Control Systems Technology, vol. 13, pp. 559-576, 2005.

[3] R. Kandiban and R. Arulmozhiyal, "Speed Control of BLDC Motor Using Adaptive Fuzzy PID Controller," Procedia Engineering, vol. 38, pp. 306-313, 2012/01/01 2012

[4] M. Kushwah and A. Patra, "Tuning PID controller for speed control of dc motor using soft computing techniques-a review," Advance in Electronic and Electric Engineering, vol. 4, pp. 141-148, 2014.

[5] W. Gubara, M. Elnaim, and S. F. Babiker, "Comparative study on the speed of DC motor using PID and FLC," in 2016 Conference of Basic Sciences and Engineering Studies (SGCAC), 2016, pp. 24-29.

[6] George Ellis, Chapter 6 - Four Types of Controllers, Editor(s): George Ellis, Control System Design Guide (Fourth Edition), Butterworth-Heinemann, 2012, Pages 97-119, ISBN 9780123859204.

[7] Joseph, Stephen Dada, Emmanuel Abidemi, Afeez Oyewola, David Mohammed, Ban. (2022). Metaheuristic Algorithms for PID Controller Parameters Tuning: Review, Approaches and Open Problems. Heliyon.

[8] Stephen Bassi Joseph, Emmanuel Gbenga Dada, Afeez Abidemi, David Opeoluwa Oyewola, Ban Mohammed Khammas. Metaheuristic algorithms for PID controller parameters tuning: review, approaches and open problems.

[9] Abushawish, Eng. Abdallah Hamadeh, Mohammed Nassif, Ali. (2020). PID Controller Gains Tuning Using Metaheuristic Optimization Methods: A survey. International Journal of Computers.

[10] Mirjalili, Seyedali Mirjalili, Seyed Lewis, Andrew. (2014). Grey Wolf Optimizer. Advances in Engineering Software.

[11] Wang, JS., Li, SX. An Improved Grey Wolf Optimizer Based on Differential Evolution and Elimination Mechanism. Sci Rep 9, 7181 (2019).

[12] Aala Kalananda, V.K.R., Komanapalli, V.L.N. A competitive learning-based Grey wolf Optimizer for engineering problems and its application to multi-layer perceptron training. Multimed Tools Appl (2023).

[13] Almufti, S.M., Ahmad, H.B., Marqas, R.B. Asaad, R.R. (2021). Grey wolf optimizer: Overview,modifications and applications. International Research Journal of Science, Technology, Education, and Management, 1(1), 44-56.

[14] Xin-She Yang, Yu-Xin Zhao, Chapter 3 - Firefly algorithm and flower pollination algorithm, Editor(s): Xin-She Yang, Nature-Inspired Computation and Swarm Intelligence, Academic Press,2020,Pages 35-48.

[15] Yang, XS. (2012). Flower Pollination Algorithm for Global Optimization. In: Durand-Lose, J., Jonoska, N. (eds) Unconventional Computation and Natural Computation. UCNC 2012. Lecture Notes in Computer Science, vol 7445. Springer, Berlin, Heidelberg.

[16] S. Lalljith, I. Fleming, U. Pillay, K. Naicker, Z. J. Naidoo and A. K. Saha, "Applications of Flower Pollination Algorithm in Electrical Power Systems: A Review," in IEEE Access, vol. 10, pp. 8924-8947, 2022.

[17] Weijia Cui, Yuzhu He, "Biological Flower Pollination Algorithm with Orthogonal Learning Strategy and Catfish Effect Mechanism for Global Optimization Problems", Mathematical Problems in Engineering, vol. 2018, Article ID 6906295, 16 pages, 2018.

[18] C. Thammarat, A. Nawikavatan and D. Puangdownreong, "Application of Flower Pollination Algorithm to PID Controller Design for Three-Tank Liquid-Level Control System," Proceedings of the 9th International Conference on Sciences, Technology and Innovation for Sustainable Well-Being (STIWB), pp.42-46, 2017.

[19] Mohamed, R., Saber, S., Askar, S. S., Abouhawwash, M. (2021). Modified Flower Pollination Algorithm for Global Optimization. Mathematics, 9(14), 1661.

[20] M. A. Hasan, R. K. Mishra and S. Singh, "Speed Control of DC Motor Using Adaptive Neuro Fuzzy Inference System Based PID Controller," 2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC), Greater Noida, India, 2019, pp. 138-143.

[21] T. Tarannum, "Intelligent Speed Control of DC Motor using ANFIS D Controller," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ICASERT.2019.8934620.

[22] Jilani, A. , Murawwat, S. and Jilani, S. (2015) Controlling Speed of DC Motor with Fuzzy Controller in Comparison with ANFIS Controller. Intelligent Control and Automation, 6, 64-74.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:10:35 UTC from IEEE Xplore. Restrictions apply.

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Evaluating the impact of wind effects on tall buildings with various plan shapes

Indrajeet Singh, Abhishek Prakash Paswan, Chitrakshi, Ritu Raj, S. Anbukumar, Rahul Kumar Meena *

Department of Civil and Environmental Engineering, Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
Aerodynamic Coefficient
Pressure
Tall Buildings
CFD
Turbulence Model

ABSTRACT

Tall building is playing important role in the growth of urbanization and has resulted in developing to the vertical rise of the structure. The data available for the designing of the tall building is not sufficient or need some modification up to very limited extent in the present scenario. In this study a rectangular shape having L: B ratio as 1.15 and octagonal plan shape are selected to investigate the wind loads on building. The area of both models is nearly equal for the comparison of the results. Pressure effects are the main concern of considered building model. When it comes to predicting wind-related events on buildings or other structures, computational fluid dynamics (CFD) is a reliable and quicker option. CFD analysis is carried out using a software program called (ANSYS-CFX). This paper presents pressure contours that are calculated at the front, back & adjacent sides of the individual tall buildings. The results clearly show that based upon the building shape, the pressures generated can be either positive pressure or negative which can be either beneficial or can have adverse effects. The result on both the building model shows that the pressure on wind ward face is more or less the same and it does not depend on the shape of the building. In the case of a symmetrical building, the leeward face will always be subjected to the suction effect, while the side faces will always exhibit an equal pressure distribution.

1. Introduction

This is well known fact that the population is migrating to urban cities for the different causes hence the requirement for the tall building is increased with effect from the past year. Because of this reason the demands of the tall building are increasing in the urban cities also present growth pattern are replacing the past low-rise buildings in to the tall buildings. Wind effects are unpredictable in the nature and varies for each model hence their investigation is must for all high-rise structure. Codal provisions and various available data is need some upgradation in various forms hence the investigation of wind generated effects is must for new building with some or minor modification. Wind effects can be identified on the tall building using wind tunnel test or either CFD [1]. The growth of modern architecture has significantly changed the pattern of wind generated effects on modern infrastructure with effect from the past conventional design of tall buildings. As the wind effects are varied for each type of building model. When it comes to climate change, one of the most important variables to consider is wind in case of the tall building as the wind load is increase as the height increases. The wind is consisting of different types of eddies of different sizes and rotational effects that are carried along in a large stream of wind moving relative to the earth surface. This is one of the most important variables to consider in the designing of high-rise structure. Because of these eddies, the wind has a gusty or turbulent characteristics [2]. Computational fluid dynamics (CFD) simulations are used extensively by different researcher in past such as [3].[4] wind effects are investigated using experimental method boundary layer wind tunnel on "+" shape of tall building. It is found that as the wind incidence angle changes on the building model wind effects are changed hence these mostly varies as per the plan cross sectional shape. The increase in the height of the building model causes the increase or decrease in the wind effects in terms of pressure or suction on the building wind ward or lee ward side [5]. The effects of wind on the different bracing system used in high rise steel structure were investigated and it is found that building load carrying capacity is affected for different pattern of bracings. Different plan cross sectional shape building model were discussed in detail by [6 7] and investigated the interference effect on the primary building model by placing equal height building model in the surrounding of the primary building model.

* Corresponding author.
E-mail address: rahul.08dtu@gmail.com (R. Kumar Meena).

https://doi.org/10.1016/j.matpr.2023.07.221

2214-7853/Copyright © 2023 Elsevier Ltd. All rights reserved. Selection and peer-review under responsibility of the scientific committee of the 2nd International Conference on Sustainable Materials and Practices for Built Environment.

Please cite this article as: Indrajeet Singh et al., *Materials Today: Proceedings*, https://doi.org/10.1016/j.matpr.2023.07.221

Due to the interference, it is found that the pressure, and the base shear is exhibited by the square plan shaped model at full blockage condition with effect from the fish plan shaped tall building model is lesser in magnitude.

Wind effects are investigated in past studies are mostly on regular type model such as square, rectangular with the variation in the length to width ratio of different cross sectional shape such as[8,9]carried out an experimental study of wind-induced pressure on tall buildings of various shapes from 0^0 to 90^0 wind incidence angle by varying the wind at 30^0 intervals and discovered that wind pressure decreases up to 60^0 wind incidence angle from the 0^0 wind incidence angle. It is also found that the "H" plan shaped building is subjected to a more pressure than the square plan shaped building. Various type of modification was provided and investigated to reduce the wind effects on tall building to make it more resistant to wind effects. [10]. Wind excitation and effects can be reduced in high-rise structures by carefully considering the construction of the different architectural feature in plan geometry of the building. Modification in the shape of a building is a mostly used option to reduce the wind effects [11 12]. Pressure characteristic in terms of wind flow and dynamic response are the earlier investigation done on different cross-sectional shape to provide most optimistic building to minimise the wind effects [13]. Presented the wind effects on "H" plan shape building [14], along with the interference of buildings for both isolated and interference condition [15] and it is observed that the in some cases interference of secondary model is reduces the wind effects on primary model [16]. Architect are now a days designing building based on some modern design tools and hence the building with modifications are useful to reduce the wind effects[17 18]. The mean Cp for octagonal building model are computed using numerical simulation and experimental result and compared the result on different sides of the tall building model.

A non-dimensional coefficient, wind pressure coefficients Cp, are used to express the wind-induced pressure at a specific building surface over a structure. Wind pressure coefficients are commonly used in the calculation of wind loads and wind- induced pressure. It is necessary to know the wind pressure coefficients in order to make accurate wind load assessments and structural designs, especially for large structures that are more susceptible to wind pressures. Full-scale measurements are difficult to carry out, and expensive also these take a long time. For the same reason, wind tunnel tests are expensive and require a significant amount of expertise. Previous work has been done in low-rise buildings with simple geometry to obtain full-scale measurements for the purpose of estimating wind-induced pressure. Wind loads in difficult structures such as high-rise buildings and non-conventional architectural structures are frequently determined through numerical analysis using Computational Fluid Dynamics (CFD) simulations.

2. Numerical simulation

The wind load is investigated on tall building is mostly done using two methods such as wind tunnel test and CFD. In the present CFD is utilized to calculate the wind effects on the equal area building model. The wind effects on a rectangular and octagon building are presented in this study. Result in various form such as mean pressure coefficient and contours are presented with graphical representation. Despite the fact that the plan area and dimensions of these two structures are nearly identical, the shape of the buildings has a significant influence on the wind behaviour. Following basic equation are used to calculate the wind effects.

Continuity equation:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \tag{1}$$

Momentum equation:

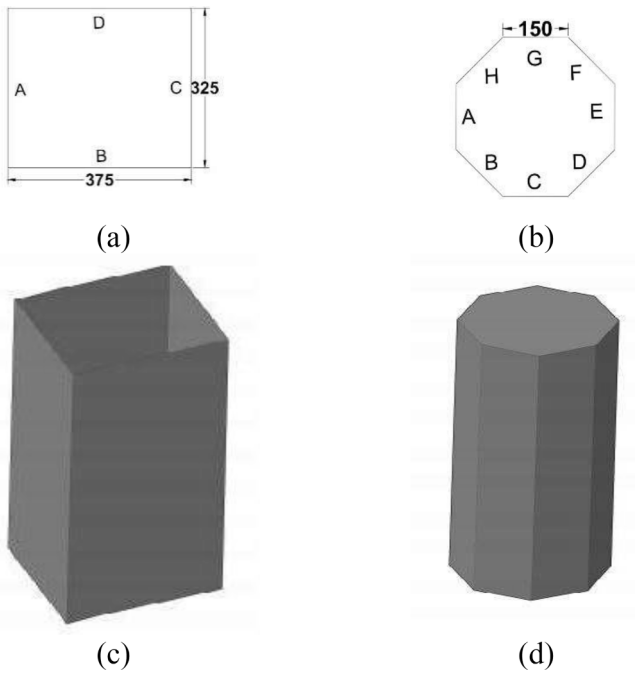


Fig. 1. Plan and isometric view of building model (a) rectangular plan (b) octagon plan (c) rectangular isometric (d) octagon isometric.

$$\rho\left(\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial y}\right)=\frac{-\partial\rho}{\partial x}+\rho g_x+\mu\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right) \tag{2}$$

Eq. (2). can be extended to y and z directions as well so as to get the entire momentum equation. The k-ε model has the benefit in modelling geometry-related factors. The connection of the conservation equations with the entire system of equations is stiff and slow convergent. In general, the Navier-Stokes equation of continuity and momentum is used in CFD. The following are the equations for the unsteady three-dimensional flow to compute the wind effects.

$$\mu_{eff}=\mu+\mu_t \tag{3}$$

Where μ_t is the turbulence viscosity. The $k-\varepsilon$ assumes the following relation between the turbulence viscosity, k and ε.

$$\frac{\partial\rho k}{\partial t}+\nabla(\rho\vec{u}k)=\nabla\left[\left(\mu+\frac{\mu_t}{\sigma_k}\right)\nabla\varepsilon\right]+\nabla u\vec{\tau}-\rho\varepsilon \tag{4}$$

2.1. Building model

Model used in this study are rectangular and octagon depicted in figure 1, the ratio of L:B is kept 1.15. Isometric view rectangular building and octagon building model is also presented. CFD output is very much depends on the knowledge of meshing, and numerical modelling. Despite the fact that simulations are very precise, they require time. Structural engineers may use parametric design to modify their models in a systematic and automated manner.

2.2. Domain and meshing

The domain boundary must be given to describe a problem with a unique solution. Unspecified boundary circumstances have a significant impact on your response. Specifying the solver, the input and exit walls, and other boundary constraints. [19] The domain, boundary conditions, and numerical simulation should all be used simultaneously, according to the proposed model and guidelines available in wind tunnel manual report no 67. A fluid domain was employed in this study to simulate a virtual wind tunnel. In ANSYS CFX, the fluid properties are given, and

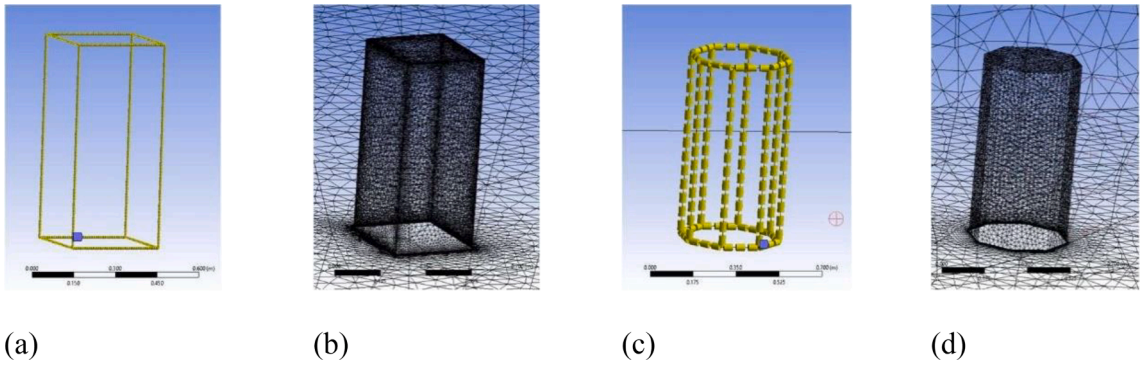


Fig. 2. Meshing (a) of edge for Model A (b) for building model A (c) of edge for Model B (d) for building model B.

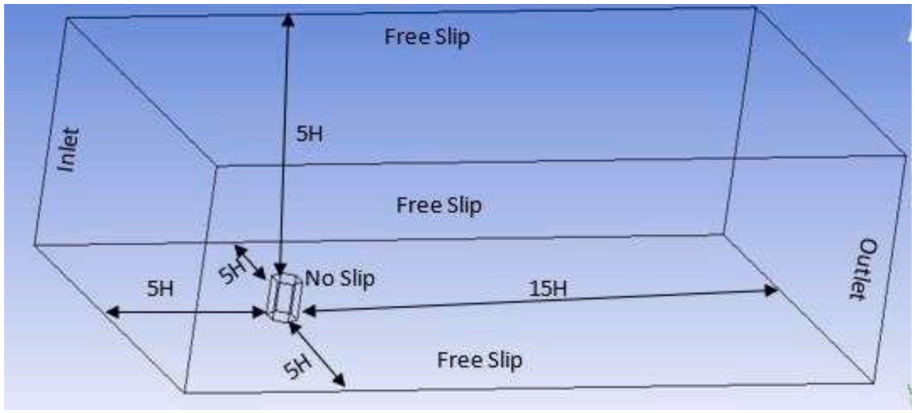


Fig. 3. Domain.

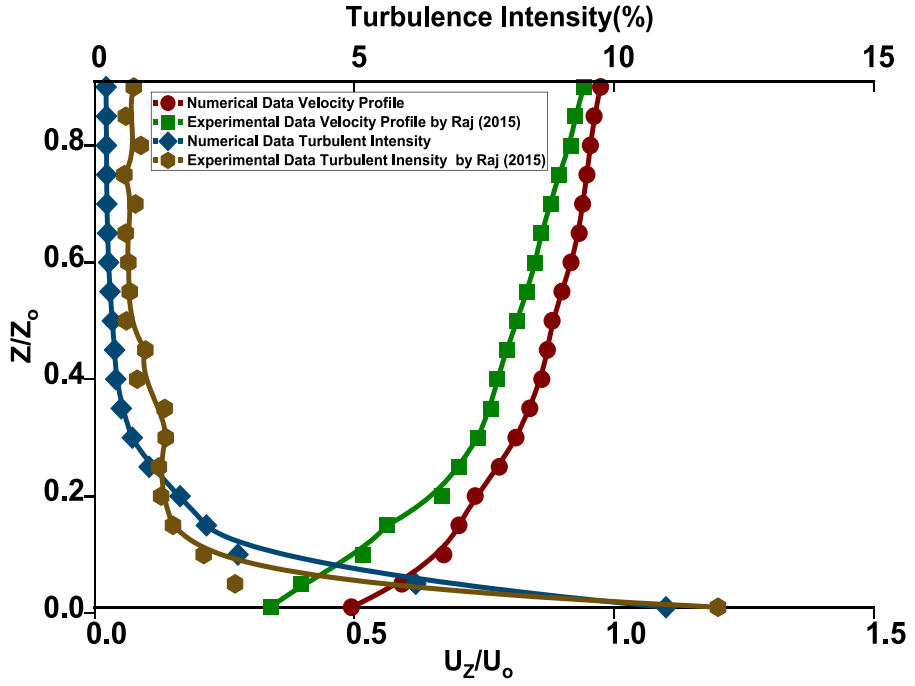


Fig. 4. Velocity profile and turbulent intensity [20].

the building sides are defined as free slip walls.

Fig. 3 depicts the domain that was used for this investigation, with the ground and model surfaces set to no slip and the side walls set to free slip. The edge meshing (a) and building meshing (b) are both finer than the domain meshing (c) when it comes to primary model. Edge meshing is shown in (c) for the octagonal building model, and building meshing is shown in (d) for the rectangular building model. A good mesh may not always result in an optimal solution, but a bad mesh will always result in

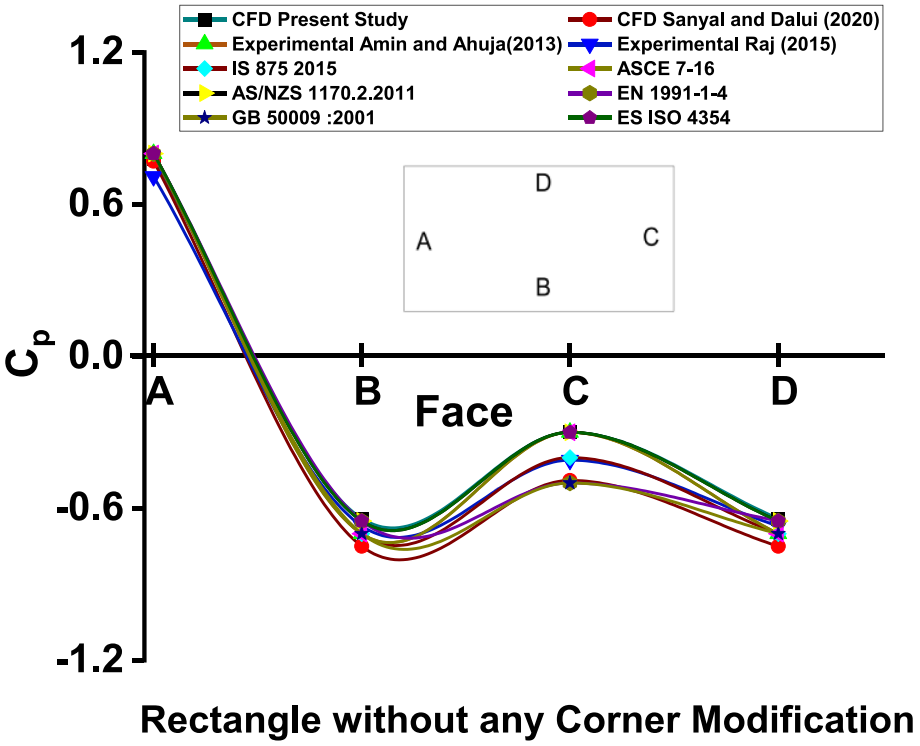


Fig. 5. Validation of rectangular model [20].

a bad solution; meshing is provided as shown in Fig. 2. Inflation is used at the interface of the tall building and the ground. It is taken into consideration in Fig. 2 that air impacts in the separation zone (b). It is difficult to provide an adequate meshing solution because the meshing is dependent on the complexity of the problem. It is impossible for the computer to solve the simulation on the actual geometric shape that was used in the simulation because the governing equation can only be applied to a specific shape and cannot be applied to any other shape. It is possible to solve governing equations in defined volumes using mesh elements and this can be done in a predictable and mathematical manner.

2.3. Velocity and turbulent intensity

In this numerical simulation, the wind input profile is imposed as a power law, with a reference velocity of 10 m/s and an alpha of 0.147 in the power law, and a reference height of 1 m all these constraints are based on the guidelines available in IS 875 part 3 and ASCE report no 67 on wind tunnel manual. Fig. 4 shows the velocity and turbulence intensity profiles.

2.4. Validation

Validation of numerical results using existing data, both numerical and experimental, is essential. [20] establishes results obtained through CFD simulations with other established simulations as well as wind load codes and experimental results for the rectangular building. The coefficient of pressure (Cp) is a dimensionless quantity that represents the ratio of the pressure at any point to the upwind pressure at roof height, and is given by following equation.

$$C_p = \frac{p - p_o}{\frac{1}{2}\rho U_Z^{\,2}} \tag{5}$$

where, p is the pressure derived on the surface of the building model by different techniques, p_o is the reference height static pressure and the air density (1.225 kg/m^3), with Uz denoting the building average wind

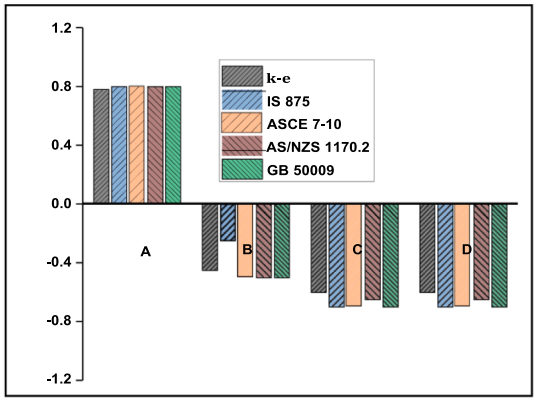


Fig. 6. Mean Cp for rectangular building model- A.

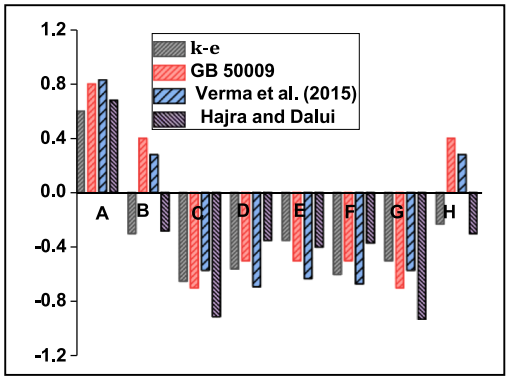


Fig. 7. Mean Cp for octagon building model-B.

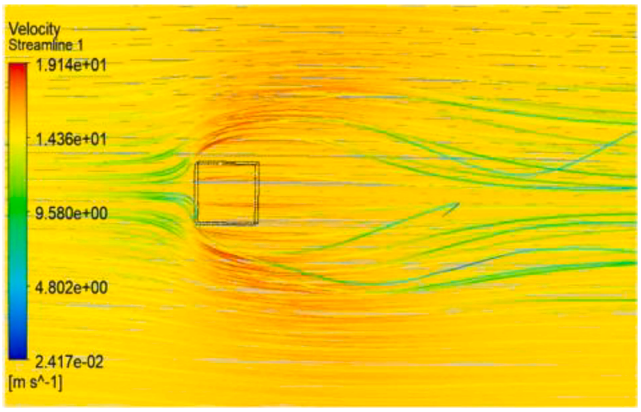


Fig. 8. Velocity streamline for rectangular model.

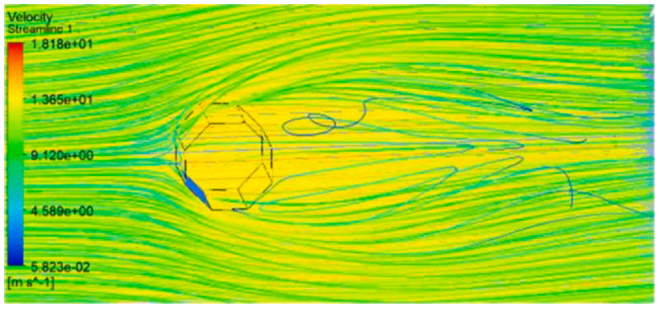


Fig. 9. Velocity streamline for rectangular model.

speed at the reference height. The results shown in fig. 5 for the present CFD study is well in agreement with the other established studies and thus the same setup has been implemented for the entire course of this study.

3. Result and discussion

3.1. Mean Cp for building model

Figures 6 and 7 compare the results of the different studies carried out on the rectangular and octagonal models respectively. The k-ε model proves to be validated when compared to the different codal provisions. However, the values of Cp of the windward and leeward faces of rectangular building differs from that of the octagonal building due to the shape and shielding effects.

3.2. Velocity stream lines

Figs. 8 and 9 depict the velocity stream line in plan shape for model A and model B respectively. The pattern represents the stream line and vortices are the main cause of creating the pressure and suction on the respective faces of the building model.

3.3. Pressure contours

The pressure contours presented below depict the distribution of pressure along the rectangular (i.e. Fig.10, faces A, B, C, D) and octagonal (i.e. Fig.11, faces A, B, C, D, E, F, G, H).

Higher pressure values are observed for the windward face of the rectangular building as compared to the octagonal. The pressure distribution is almost uniform for the side faces of both the building model considered in the study. However, a greater suction pressure is observed for the rectangular building in the leeward side due to its higher shielding in compression to the octagonal shape model.

4. Conclusion

Several conclusions can be drawn from the research that has been conducted to investigate the wind effects. These are based on the CFD analysis done on ANSYS CFx along with the comparisons with experimental data. Some conclusions based on the investigation on equal area building model are as following.

- Regular plan form buildings exhibit symmetrical pressure distributions between symmetrical faces of rectangular & octagonal buildings.
- Pressure contours on the front face of rectangular building shows high pressure oin comparison to the side faces of the building model. While the suction is observed in the lee ward side of the model.
- The octagonal shaped building showed concentration of pressure on the front face which is slightly increased with effect to the rectangular mode as wind effects are increased with the height of the building.
- The adjacent faces with respect to the wind ward face shows the higher magnitude to lower magnitude if moving along the downstream of wind but vertically it remained almost constant.
- Faces C & D represent slight negative pressure distribution on the adjoining edges to the faces. Faces G & F are in the less pressure as compared to other faces with slight increase with the height of the building in the case of the octagonal building model.
- In order to forecast wind-related events on buildings and other various types of structures, Computational Fluid Dynamics (CFD) may be used as a alternative method to visualization of wind effects.

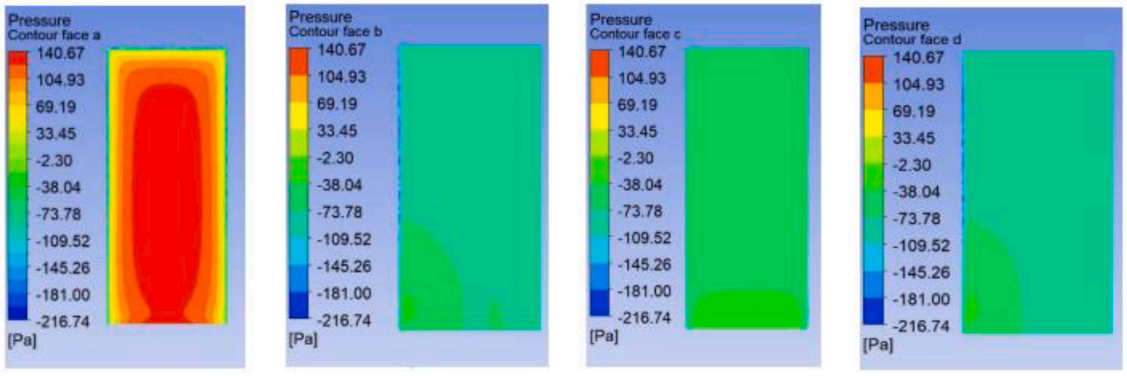


Fig. 10. Pressure Contours for rectangular building model.

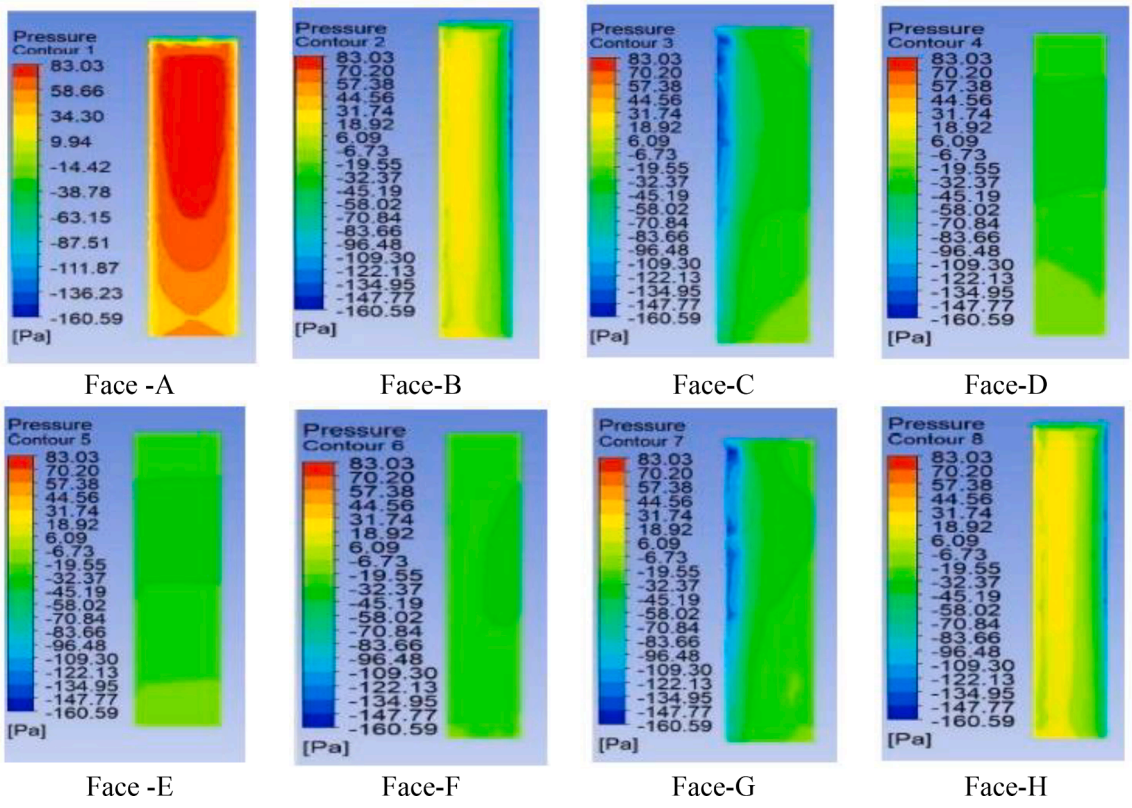


Fig. 11. Pressure Contours for rectangular building model.

CRediT authorship contribution statement

Indrajeet Singh: Writing – review & editing. **Abhishek Prakash Paswan:** Validation. **Chitrakshi:** Data curation. **Ritu Raj:** Resources, Project administration, Funding acquisition. **S. Anbukumar:** Resources, Project administration, Funding acquisition. **Rahul Kumar Meena:** Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition.

Funding

The authors Indrajeet, Abhishek, Chitrakshi and Rahul thanks the Delhi Technological University, Delhi Civil and Environmental Engineering Department for providing the institutional fellowship for PhD.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

The authors are thankful to the faculty of civil engineering department of the university for providing the research lab and computational resources to support this study.

References

[1] R. Kar, S.K. Dalui, Wind interference effect on an octagonal plan shaped tall building due to square plan shaped tall buildings, Int. J. Adv. Struct. Eng. 8 (2016) 73–86, https://doi.org/10.1007/s40091-016-0115-z.
[2] Hima Chandan, D., Pradeep Kumar, R.: Numerical Simulation of Wind Analysis of Tall Buildings Computational Fluid Dynamics Approach. (2014).
[3] K.B. Rajasekarababu, G. Vinayagamurthy, S. Selvi Rajan, Experimental and computational investigation of outdoor wind flow around a setback building, Build. Simul. 12 (2019) 891–904, https://doi.org/10.1007/s12273-019-0514-8.
[4] R. Raj, A.K. Ahuja, Wind loads on cross shape tall buildings, J. Acad. Industrial Res. (JAIR). 2 (2013) 111–113.
[5] Y.C. Kim, Y. Tamura, H. Tanaka, K. Ohtake, E.K. Bandi, A. Yoshida, Wind-induced responses of super-tall buildings with various atypical building shapes, J. Wind Eng. Ind. Aerodyn. 133 (2014) 191–199, https://doi.org/10.1016/j.jweia.2014.06.004.
[6] Meena, R.K., Awadhiya, G.P., Paswan, A.P., Jayant, H.K.: Effects of Bracing System on Multistoryed Steel Building. IOP Conference Series: Materials Science and Engineering. 1128, 012017 (2021). https://doi.org/10.1088/1757-899X/1128/1/012017.
[7] Pal, S., Raj, R., Anbukumar, S.: Comparative study of wind induced mutual interference effects on square and fish-plan shape tall buildings. Sādhanā. 0123456789, (2021). https://doi.org/10.1007/s12046-021-01592-6.
[8] N.S. Fouad, G.H. Mahmoud, N.E. Nasr, Comparative study of international codes wind loads and CFD results for low rise buildings, Alex. Eng. J. 57 (2018) 3623–3639, https://doi.org/10.1016/j.aej.2017.11.023.
[9] S.K. Nagar, R. Raj, N. Dev, Experimental study of wind - induced pressures on tall buildings of different shapes, Wind Struct. Int. J. 5 (2020) 441–453.
[10] M. Mallick, A. Mohanta, A. Kumar, V. Raj, Modelling of wind pressure coefficients on C- shaped building models, Model. Simul. Eng. 2018 (2018), https://doi.org/10.1155/2018/6524945.
[11] M.S. Thordal, J.C. Bennetsen, S. Capra, H.H.H. Koss, Towards a standard CFD setup for wind load assessment of high-rise buildings: Part 1 – Benchmark of the CAARC building, J. Wind Eng. Ind. Aerodyn. 205 (2020), 104283, https://doi.org/10.1016/j.jweia.2020.104283.
[12] J. Allegrini, B. Lopez, The influence of angular configuration of two buildings on the local wind climate, J. Wind Eng. Ind. Aerodyn. 156 (2016) 50–61, https://doi.org/10.1016/j.jweia.2016.07.008.
[13] Dalui, S.K., Kar, R., Hajra, S.: Interference Effects on Octagonal Plan Shaped Tall building under Wind - A Case Study. 9, 74–81 (2015).
[14] R. Raj, T. Rana, T. Anchalia, U. Khola, Numerical study of wind excited action on H Plan- shaped tall building, Int. J. Emerg. Technol. 11 (2020) 591–605.

[15] R. Paul, S.K. Dalui, Wind effects on 'Z' plan-shaped tall building: a case study, Int. J. Adv. Struct. Eng. 8 (2016) 319–335, https://doi.org/10.1007/s40091- 016-0134-9.
[16] Liaw, K.F.: Simulation of Flow around Bluff Bodies and Bridge Deck Sections using CFD. 268 (2005).
[17] D.S.K. Verma, A. Roy, S. Lather, M. Sood, CFD simulation for wind load on octagonal tall buildings, Int. J. Eng. Trends Technol. 24 (2015) 211–216, https://doi.org/10.14445/22315381/ijett-v24p239.
[18] S. Hajra, S.K. Dalui, Numerical investigation of interference effect on octagonal plan shaped tall buildings, Jordan J. Civil Eng. 10 (2016) 462–479.
[19] Revuz, J.: Revuz , Julia (2011) Numerical simulation of the wind flow around a tall building and its dynamic response to wind excitation . PhD thesis , University of Nottingham . Numerical around simulation of the wind flow and its dynamic a tall building. (2011).
[20] R.K. Meena, R. Raj, S. Anbukumar, Wind excited action around tall building having different corner configurations, Adv. Civil Eng. 2022 (2022) 1–17, https://doi.org/10.1155/2022/1529416.

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 27 2023

Evaluation of various beams using computational tools FREE

Bhagya Jayant; Rahul Kumar Meena; Aastha Singh; Agyeya Mishra; Navneet Dalal; Rajvir Singh; Shailendra Kumar; Ritu Raj

Check for updates

AIP Conference Proceedings 2721, 020042 (2023)
https://doi.org/10.1063/5.0153989

View Online
Export Citation
CrossMark

10 August 2023 06:53:53

500 kHz or 8.5 GHz?
And all the ranges in between.
Lock-in Amplifiers for your periodic signal measurements

Find out more
Zurich Instruments

AIP Publishing

Evaluation of Various Beams Using Computational Tools

Bhagya Jayant[1] , Rahul Kumar Meena[1] , Aastha Singh[1] , Agyeya Mishra[1], Navneet Dalal[1] , Rajvir Singh[1] , Shailendra Kumar[1] , Ritu Raj [1,a]

[1]*Delhi Technological University, Delhi, India,*

[a)] *Corresponding Author: rituraj@dtu.ac.in*

Abstract. This study aims to demonstrate a Python- based model or Python scripts for static analysis of beams using force method for a set of beams with changing cross – sections categorized on the basis of types of beam structure and loading conditions. It aims to fill the gap of less research in the domain of static analysis of beams by formulating the beam analysis into an algorithm using the programming language, Python. The types of beam structures under consideration are cantilever, propped cantilever, simply supported, and fixed beam. And the loading conditions of point load and uniformly distributed load (UDL) are taken as per subject logic. The entire program has been prepared and verified by referring to standard books of Civil Engineering. Readers may run the provided Python script or code in any Python release for Windows and/or Pythonista. The research paper targets to simplify the concepts in this domain in a fast, an efficient, a user – friendly, and an accurate way. Fellow researchers working in a similar field of interest may utilize this paper as an aiding tool for their research and save their time and manual effort.

Keywords. Structural analysis; Various Type of Beams; Force Method; Python; Deflection.

10 August 2023 06:53:53

INTRODUCTION

Civil engineering structures have a series of components like beams, columns, slabs, arches, etc. All these components are designed to take the load of the superstructure and distribute it through various suitable foundations to sub-soil. The structural components are designed and need to be checked for deflection and slope. This is a check. If these checks are made manually, the possibility of human error is increasing. That is why the development of data algorithms and PYTHON programming is used to obtain this deflection and slope. All the software which analyses the civil engineering structure mainly uses finite element techniques. Some researchers used ANSYS to find the response of the tall structure (N. Gaur, R. Raj, and P. K. Goyal, 2021), (A. Kumar and R. Raj, 2021), (N. Gaur and R. Raj, 2021), (R. K. Meena, R. Raj, and S. Anbukumar, 2021) and predicted the effect of wind to design the high rise projects. The response of the structure in terms of base shear and deflection is also obtained by various studies (R. K. MEENA, G. P. Awadhiya, A. P. Paswan, and H. K. Jayant), (R. Raj and A. K. Ahuja, 2013), (S. Mahajan, V. Yadav, R. Raj, and R. Raj), (R. Raj, A. Sharma, and S. Chauhan, 2018). The majority of the computational methods make use of the fundamentals of data algorithms (C. A. Shaffer and V. Tech, 2011), (A. Neubauer, J. Freudenberger, and V. Kühn, 2007), (Chapter 1 Pointers , Arrays , and Structures., 1996), (Shaw, 2017), (Johansson, 2018). Before the start of this study, a large amount of research was done, and the major outcomes are explained in various sections of this study. The computational method which is used to get the deflection and slope has a lower degree of variation. That is why the development of some data-programming techniques is required and this study presents the same. The basics of structural analysis are found and studied in detail (Karnovsky, 2012), (J. J. Connor and S. Faraji, 2013), (Hibbeler, 2012), (B. G. Nielson and J. C. Mccormac, 2017), (K. Derucher, C. Putcha, and U. Kim, 2013). The various sections of this study need the basics of various structural components, and the same is obtained from the different standards books. These various sections explain some major conclusions. Lee et al. (B. Koo Lee, J. F. Wilson, and S. Jin Oh, 1993) used the Runge-Kutta and Regula-Falsi methods to solve the elastica of cantilever beams with variable cross sections numerically and experimentally. Eci et al. (M. C. Ece, M. Aydogdu, and V. Taskin, 2007) explained vibration mechanics of anisotropic beam with a variable cross section were calculated numerically. Non-uniformity in the cross-sectional area influences the natural frequencies and the model shape. The amplitude of vibration is found to be increased if there is an increment in the cross section, while there will be a decrement in the amplitude of vibration if the cross-sectional area of the beam is reduced. Sapountzakis and Tsiatas (E. J. Sapountzakis and G. C. Tsiatas, 2007), (E. J. Sapountzakis and G. C. Tsiatas, 2005) studied the boundary element method for analysis of the elastic flexural

2nd International Conference on Futuristic and Sustainable Aspects in Engineering and Technology
AIP Conf. Proc. 2721, 020042-1–020042-27; https://doi.org/10.1063/5.0153989
Published by AIP Publishing. 978-0-7354-4576-5/$30.00

buckling of composite beams with variable cross sections. Rahman and Kowser (M. A. Rahman and M. A. Kowser, 2007) performed nonlinear analysis of cantilever shape memory alloy beams of variable cross sections and found that the reduction of more material can be done in the beam which has the unform strength and variable cross section. Karnovsky (Karnovsky, Theory of Arched Structures, 2012) explains the strength, stability, and vibration involved in the arched shape structures. Boiangiu et al. (M. Boiangiu, 2014) performed vibration analysis on numerous structures with uniform and variable cross sections, and verified the transfer matrix to calculate the natural frequencies involved in multi-step beams. Rojas and Ramirez (A. L. Rojas and J. M. Ramirez, 2014) developed a mathematical model that is subjected to uniformly distributed load, which has a variable cross section. Using this method, the moment and displacement at any cross section of the beam can be obtained mathematically. Dumitrache (Dumitrache, 2012) Shear stress in a beam with a variable cross section that is subjected to uniformly distributed load. Sapountzakis and Panagos (E. J. Sapountzakis and D. G. Panagos, 2008) presented the analogue equation method that is employed for the nonlinear analysis of beams with simple or multiple connected variable cross sections and found the shear deformations etc. The study findings are utilized to confirm the structure's integrity. As a result, structural analysis is an important element of structural engineering. All structures may be divided into two categories: statically determinate structures and statically indeterminate structures.

The force technique is a method for evaluating statically indeterminate structures using force values as fundamental variables. Establishing the displacement restrictions, also known as the geometric compatibility equations, is a crucial step. This approach involves creating a set of compatibility equations based on the number of redundant forces in the structure and concurrently solving these equations to estimate the size of the redundant forces. Once the redundant forces are identified, the structure becomes determinate and can be fully studied using equilibrium conditions.

As a result of various other qualitative secondary research performed, it is observed that stability analysis and vibration analysis are better interpreted using computer languages like Python, C, and C++ along with modeling software recently. This paper is another such piece of research primarily focusing on computerizing the static analysis of beams. The primary goal of this paper is to fill the research gap of less research in the domain of structural analysis by providing easy-to-use Python scripts for analyzing types of beams with changing cross-sections under various loading situations.

10 August 2023 06:53:53

METHODOLOGY

In general terms, statically determinate beams are those beams whose reaction forces may be determined using the equilibrium equations, in $\Sigma F_X = 0$, $\Sigma F_Y = 0$, $\Sigma F_Z = 0$ and $\Sigma M = 0$ forms. And indeterminate beams are those beams that require compatibility equations in addition to the equilibrium equations to determine all of their (unknown) force reactions.

Layout

The following flowchart demonstrates the categorization of beams into different cases based on types of beam structure (Cantilever, Propped cantilever, simply supported and fixed beam) and loading condition (Point load, mid – way to span length, UDL over the entire span length).

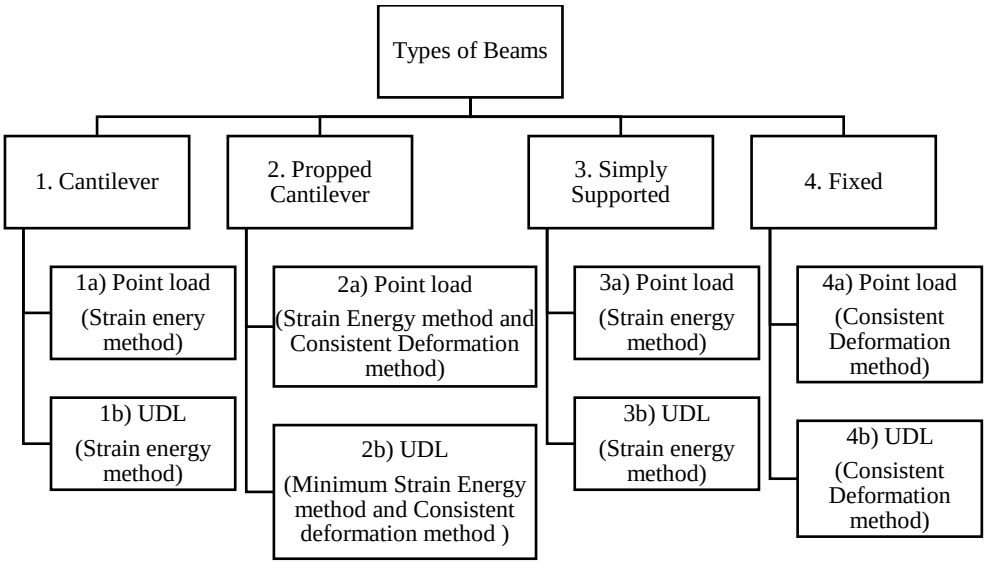


Fig 1. Flowchart depicting cases (types of beams with their loading conditions) & force method for analysis

In this paper, use of the Castigliano's theorem, Strain Energy method, and the Consistent Deformation method is done for the analysis of different beam cases.

Force Method for Analysis of Beams Used in Program Code

Castigliano's theorem

Castigliano's first theorem – For beams, trusses and frames, if the material is linearly elastic, supports are unyielding and temperature is constant, then the first partial derivative of total strain energy with respect to deflection or slope will give the load or moment at that point in the direction corresponding to that deflection component.

$$\frac{\partial U}{\partial \Delta} = P \tag{1}$$

$$\frac{\partial U}{\partial \theta} = M \tag{2}$$

Where U = strain energy, P = load, Δ = Deflection, M = Moment

Castigliano's second theorem – At constant temperature when the materials are linearly elastic and supports are unyielding, the first partial derivative of total strain energy with respect to load or moment will give the deflection or slope respectively. This theorem is also called the theorem of least work.

$$\frac{\partial U}{\partial P} = \Delta \tag{3}$$

$$\frac{\partial U}{\partial M} = \theta \tag{4}$$

10 August 2023 06:53:53

Strain energy method

When a gradual load is applied on an elastic body, the body is strained and work is done on the body which is stored in the form of internal energy. This internal work done by the body in resisting the straining is called strain energy. Mathematically, it may be expressed as –

(a) For axial load in a uniform body

$$\boldsymbol{U = \left(\frac{1}{2}\right) \times \Delta L \times P} \tag{5}$$

Which may be reduced to $U = \left(\frac{P^2.L}{2A.E}\right)$ and $U = \int \left(\frac{P^2}{2AE}\right).dx$ and is generally used in truss.

(b) For bending moment (generally used in frames)

$$\boldsymbol{U = \int \left(\frac{M_x^2}{2EI}\right).dx} \tag{6}$$

Where U = strain energy, P = Load, Δ = Deflection, AE = Axial rigidity, M = Moment

Minimum strain Energy method – If the strain energy in an elastic body is U and R_1, R_2 etc. are the redundant reactions, then if there are no support movements and no change in temperature, then the redundant reactions R_1, R_2 etc. must be such that the strain energy is minimum.

$$\boldsymbol{\frac{\partial U}{\partial Ri} \rightarrow 0} \tag{7}$$

Consistent deformation method

It is another force method used for analyzing structures using equilibrium and compatibility conditions to determine unknown in statically indeterminate structures. The basic steps are as follows.

1. Determine the degree of static indeterminacy.
2. Eliminate the redundant reaction and find out the displacement in its direction.
3. Eliminate the applied load and find out the displacement in its direction.
4. Obtain compatibility equation and solve for redundant reaction.

10 August 2023 06:53:53

CALCULATIONS

The table displays the results gotten after manual calculations in beam cases using different methods.

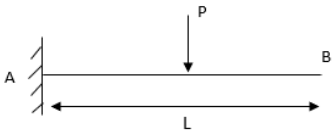
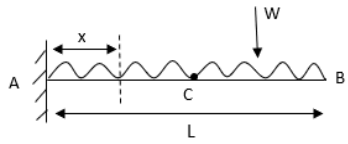
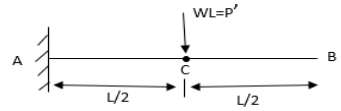
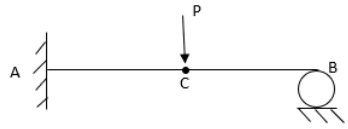
Case	Type of Beam	Loading Condition	Figure	Method used for analysis	Analysis
1a	Cantilever	Point load mid – way to span length		Strain Energy Method	1. Strain Energy, $U = \frac{P^2L^3}{6EI}$ 2. Deflection at B, $\delta_B = \frac{PL^3}{3EI}$
1b	Cantilever	UDL throughout span length	 	Strain Energy Method	1. Strain Energy, $U = \frac{W^2L^4}{8EI}$ 2. Deflection at B, $\delta_B = \frac{5WL^4}{48EI}$
2a	Propped Cantilever	Point load mid – way to span length		Strain Energy Method	Redundant Reaction $R = \frac{5P}{14}$

TABLE 1. Categorization of beams into cases

10 August 2023 06:53:53

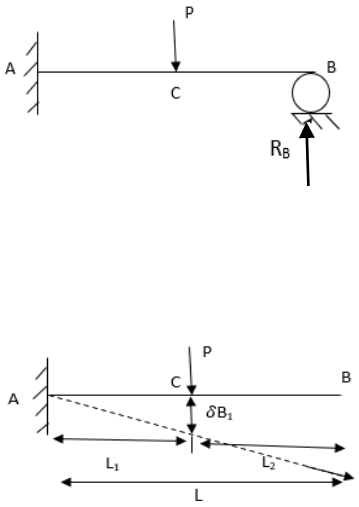
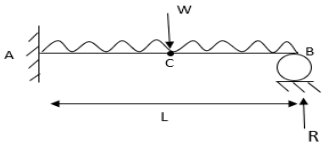
Case	Type of Beam	Loading Condition	Figure	Method used for analysis	Analysis
2a		Point load mid – way to span length		Consistent Deformation Method	1. Redundant Reaction at B, $R_B = -\left(\frac{PL_1^3}{3EI} + \left(\frac{PL_1^2}{2EI}\right)L_2\right)$ 2. Redundant Reaction at A, $R_A = P - R_B$ 3. Moment at A $M_A = R_B L - PL_1$
2b	Propped Cantilever	UDL throughout span length		Strain Energy Method	Redundant Reaction $R = \frac{3}{8}WL$
				Consistent Deformation Method	1. Redundant Reaction at B, $R_B = -\frac{3}{8}WL$ 2. Redundant Reaction at A, $R_A = WL - R_B$ 3. Moment at A $M_A = R_B(L) - \frac{WL^2}{2}$

TABLE 1 cont. Categorization of beams into cases

10 August 2023 06:53:53

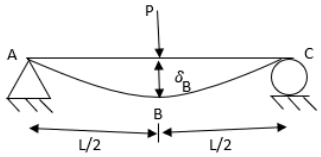
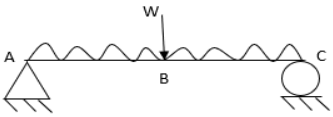
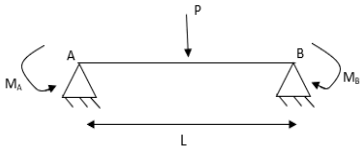
Case	Type of Beam	Loading Condition	Figure	Method used for analysis	Analysis
3a	Simply Supported	Point load mid – way to span length		Strain Energy Method	1. Strain Energy, $U = \frac{P^2 L^3}{96EI}$ 2. Deflection at B, $\delta_B = \frac{2PL^3}{96EI}$
3b	Simply Supported	UDL throughout span length		Strain Energy Method	1. Strain Energy, $U = \frac{P^2 L^3}{96EI}$ 2. Deflection at B, $\delta_B = \frac{WL^4}{48EI}$.
4a	Fixed	Point load mid – way to span length		Consistent Deformation Method	1. Moment at A $M_A = \frac{PL}{8}$ 2. Moment at B $M_B = \frac{PL}{8}$

TABLE 1 cont. Categorization of beams into cases

10 August 2023 06:53:53

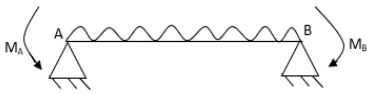
Case	Type of Beam	Loading Condition	Figure	Method used for analysis	Analysis
4b	Fixed	UDL throughout span length		Consistent Deformation Method	1. Moment at A $M_A = \frac{WL^2}{12}$ 2. Moment at B $M_B = \frac{WL^2}{12}$

TABLE 1 cont. Categorization of beams into cases

PYTHON CODE

```
print("\t\t\t\t\tWELCOME TO THE COMPUTERISED VERSION OF STRUCTURAL ANALYSIS OF BEAMS")
i = 0
while (i < 148):
    print("*", end='')
    i = i + 1

# Printing the initial instructions for our structural analysis program
print(
    "\n \t Instructions:- \n \t 1:-> Select the type of beam to analyse. \n \t 2:-> Specify the loading condition. \n \t 3:-> Select the preferred method of analysis. \n \t 4:-> Enter the corresponding values.\n")

j = 0
while (j < 148):
    print("*", end='')
    j = j + 1

# User is asked whether he or she wishes to begin the beam analysis program or not
choice = input("\nPlease enter 'y' to begin and 'n' to discontinue : ")

# If the user wishes to begin the program for beam analysis then the following while loop will be executed
while (choice == 'y'):

    # Beam menu starts here
    print("\nThe various types of beams available under Force method of analysis are:-")
```

10 August 2023 06:53:53

```
print("1:-> Cantilever Beam")
print("2:-> Propped Cantilever Beam")
print("3:-> Simply Supported Beam")
print("4:-> Fixed Beam")

# User is asked to select the type of beam he or she wishes to analyse
type_of_beam = input("\nPlease select the type of beam to analyse (1, 2, 3 or 4) : ")

# If the type of beam selected is cantilever beam then the following if statement will be executed
if type_of_beam == '1':
    print("You have selected Cantilever Beam.")

    # Loading conditions menu starts here
    print("\nThe various types of loading conditions available for cantilever beam are : ")

    # Point Load for cantilever beam
    print("\n1:-> Point Load :- ")
    print("                        |P")
    print("                        |")
    print("                        V")
    print("      |----------------------")
    print("      A                   B")
    print("\n")

    # Uniformly Distributed Load for cantilever beam
    print("2:-> Uniformly Distributed Load :- ")
    print("               |W")
    print("               |")
    print("               V")
    print("       /\/\/\/\/\/\/\/\/\/\/\/\")
    print("      |----------------------")
    print("      A                   B")
    print("\n")

    # User is asked to select the type of loading condition he or she wishes to apply to the beam
    load_cantilever = input("Please specify the loading condition (1 or 2) : ")

    # If the type of loading condition selected is point load then the following if statement will be executed
```

10 August 2023 06:53:53

```
if load_cantilever == '1':
    print("You have selected Point Load as the loading condition.")

    # User is asked to enter values for load, length of beam, modulus of elasticity and moment of inertia
    P = input("\nEnter load magnitude (P) in N : ")
    L = input("Enter length of beam (L) in mm : ")
    E = input("Enter modulus of elasticity (E) of the material of beam in N/mm^2 : ")
    I = input("Enter moment of inertia (I) of beam in mm^4 : ")

    # Menu for various structural analysis methods starts here
    print("\nThe various methods of analysis available are : ")
    print("1:-> Strain Energy Method")

    # User is asked to select the method of structural analysis
    analysis_cantilever = input("\nPlease select the method of analysis : ")

    # If the method of analysis selected is Strain Energy Method then the following if statement will be executed
    if analysis_cantilever == '1':
        print("You have selected Strain Energy Method.")

        print("\n\n\n\nSTEP 01 : CALCULATING STRAIN ENERGY (u)")
        u = (float(P) * float(P) * float(L) * float(L) * float(L)) / (6 * float(E) * float(I))
        print("\nStrain energy in this case, represented by u, is given as = (P^2*L^3)/(6*E*I) = " + str(
            u * 0.001) + " N-m")

        print("\n\nSTEP 02 : FINDING DEFLECTION AT POINT 'B'")
        deflection = (float(P) * float(L) * float(L) * float(L)) / (3 * float(E) * float(I))
        print("\nDeflection at point 'B' is given as = (PL^3)/(3*E*I) = " + str(deflection * 0.001) + " m")

        # User is asked whether he or she wishes to continue the analysis or not
        choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")

# If the type of loading condition selected is uniformly distributed load then the following elif statement will be executed
else:
    print("You have selected Uniformly Distributed Load as the loading condition.")

    # User is asked to enter values for uniformly distributed load, length of beam, modulus of elasticity and moment of inertia
    W = input("\nEnter uniformly distributed load (w) in N/mm : ")
```

10 August 2023 06:53:53

```
    L = input("Enter length of beam (L) in mm : ")
    E = input("Enter modulus of elasticity (E) of the material of beam in N/mm^2 : ")
    I = input("Enter moment of inertia (I) of beam in mm^4 : ")

    # Menu for various structural analysis methods starts here
    print("\nThe various methods of analysis available are : ")
    print("1:-> Strain Energy Method")

    # User is asked to select the method of structural analysis
    analysis_cantilever = input("\nPlease select the method of analysis : ")

    # If the method of analysis selected is Strain Energy Method then the following if statement will be executed
    if analysis_cantilever == '1':
        print("You have selected Strain Energy Method.")

        print('\n\n\n\nSTEP 01 : CALCULATING STRAIN ENERGY (u)')
        u = (float(W) * float(W) * float(L) * float(L) * float(L) * float(L)) / (8 * float(E) * float(I))
        print("\nStrain energy in this case, represented by u, is given as = (W^2*L^4)/(8*E*I) = " + str(
            u * 0.001) + " N-m")

        print("\n\nSTEP 02 : FINDING DEFLECTION AT POINT 'B'")
        deflection = (5 * float(W) * float(L) * float(L) * float(L) * float(L)) / (48 * float(E) * float(I))
        print("\nDeflection at point 'B' is given as = (5*W*L^4)/(48*E*I) = " + str(deflection * 0.001) + " m")

        # User is asked whether he or she wishes to continue the analysis or not
        choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")


# If the type of beam selected is propped cantilever beam then the following if statement will be executed
elif type_of_beam == '2':
    print("You have selected Propped Cantilever Beam.")

    # Loading conditions menu starts here
    print("\nThe various types of loading conditions available for propped cantilever beam are : ")

    # Point Load for propped cantilever beam
    print("\n1:-> Point Load :- ")
    print("                |P")
```

10 August 2023 06:53:53

```
print("                |")
print("                V")
print("      |----------------------")
print("                          O")
print("      A         B         C")
print("\n")

# Uniformly Distributed Load for propped cantilever beam
print("2:-> Uniformly Distributed Load :- ")
print("                |W")
print("                |")
print("                V")
print("       /\/\/\/\/\/\/\/\/\/\/\/\/\/\")
print("      |----------------------")
print("                          O")
print("      A         B         C")
print("\n")

# User is asked to select the type of loading condition he or she wishes to apply to the beam
load_proppedcantilever = input("Please specify the loading condition (1 or 2) : ")

# If the loading condition selected is Point Load then the following if statement will be executed
if load_proppedcantilever == '1':
    print("You have selected Point Load as the loading condition.")

    # User is asked to enter values for load, length of beam, modulus of elasticity and moment of inertia
    P = input("\nEnter load magnitude (P) in N : ")
    L = input("Enter length of beam (L) in mm : ")
    E = input("Enter modulus of elasticity (E) of the material of beam in N/mm^2 : ")
    I = input("Enter moment of inertia (I) of beam in mm^4 : ")

    # Menu for various structural analysis methods starts here
    print("\nThe various methods of analysis available are : ")
    print("1:-> Minimum Potential Energy Theorem Method")
    print("2:-> Consistent Deformation Theorem Method")

    # User is asked to select the method of structural analysis
    analysis_proppedcantilever = input("\nPlease select the method of analysis (1 or 2) : ")
```

10 August 2023 06:53:53

```
    # If the method of analysis selected is Minimum Energy Theorem Method then the following elif statement will be executed
    if analysis_proppedcantilever == '1':
        print("You have selected Minimum Potentital Energy Theorem Method.")

        print(
            "\n\n\n\nApproach:- We will first find redundant for the two the spans AB and BC and then we will superimpose them to find the redundant for the complete beam. ")
        redundantAB = (float(P) * 5) / 14
        print("\nRedundant reaction for span AB = R\' = (5*P)/14 = " + str(redundantAB) + " N")
        print("Redundant reaction for span BC = R\" = 0 N")
        print("\nTherefore, total redundant reaction for beam AC = R = R\' + R\" = (5*P)/14 = " + str(
            redundantAB) + " N")

        # User is asked whether he or she wishes to continue the analysis or not
        choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")

    # If the method of analysis selected is Consistent Deformation Theorem Method then the following elif statement will be executed
    elif analysis_proppedcantilever == '2':
        print("You have selected Consistent Deformation Theorem Method.")

        print("\n\n\n\nSTEP 01: CALCULATING DEGREE OF STATIC INDETERMINACY (Ds) ")
        print("\nNumber of actual reactions, R = 3")
        print("Number of equations available, E = 2")
        print("Degree of Static Indeterminacy, Ds = R - E = 1")

        print("\n\nSTEP 02: REMOVING REDUNDANT REACTIONS AND CALCULATING DISPLACEMENT DUE TO LOADING")
        print("\nLet 'Rb' be the redundant reaction at B.")
        deltaB1 = -(5 * float(P) * float(L) * float(L) * float(L)) / (48 * float(E) * float(I))
        print("Displacement at B due to applied loading, Delta(B1) = -(5*P*L^3)/(48*E*I) = " + str(
            deltaB1 * 0.001) + " m (Taking clockwise direction as negative)")

        print("\n\nSTEP 03: REMOVING LOADING AND CALCULATING DISPLACEMENT DUE TO REDUNDANT REACTIONS")
        Rb = (5 * float(P) / 16)
        deltaB2 = (float(Rb) * float(L) * float(L) * float(L)) / (3 * float(E) * float(I))
        print("\nDisplacement at B due to redundant reaction at B, Delta(B2) = (Rb*L^3)/(3*E*I) = " + str(
            deltaB2 * 0.001) + " m")
```

10 August 2023 06:53:53

```
        print("\n\nSTEP 04: OBTAINING COMPATIBILITY EQUATION AND SOLVING FOR 'Ma' and 'Mb'")
        print(
            "\nApproach:- To solve for redundant reaction, we need one more equation which can be obtained by compatability. It is quite evident that the displacement at B is zero.")
        print("Total Displacement at B, Delta(B) = Delta(B1) + Delta(B2) = 0")
        print(
            "Substituting the values and solving the above equation, we get redundant reaction at B, Rb = (5*P)/16 = " + str(
                Rb) + " N")
        Ra = (11 * float(P)) / 16
        Ma = -(3 * float(P) * float(L) * float(L)) / 16
        print("\nUsing the value of redundant reaction at B 'Rb', we can also calculate the following:")
        print("Reaction at A, 'Ra' = P - Rb = P - (5*P)/16 = (11*P)/16 = " + str(Ra) + " N")
        print("Bending Moment at A, 'Ma' = Rb - (P*L/2) = -(3*P*L^2)/16 = " + str(
            Ma * 0.001) + " N-m (Taking clockwise direction as negative)")

        # User is asked whether he or she wishes to continue the analysis or not
        choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")

    # If the type of loading condition selected is uniformly distributed load then the following elif statement will be executed
    else:
        print("You have selected Uniformly Distributed Load as the loading condition.")

        # User is asked to enter values for uniformly distributed load, length of beam, modulus of elasticity and moment of inertia
        W = input("\nEnter uniformly distributed load (w) in N/mm : ")
        L = input("Enter length of beam (L) in mm : ")
        E = input("Enter modulus of elasticity (E) of the material of beam in N/mm^2 : ")
        I = input("Enter moment of inertia (I) of beam in mm^4 : ")

        # Menu for various structural analysis methods starts here
        print("\nThe various methods of analysis available are : ")
        print("1:-> Minimum Potential Energy Theorem Method")
        print("2:-> Consistent Deformation Theorem Method")

        # User is asked to select the method of structural analysis
        analysis_proppedcantilever = input("\nPlease select the method of analysis (1 or 2) : ")

        # If the method of analysis selected is Minimum Potential Energy Method then the following elif statement will be executed
```

10 August 2023 06:53:53

```
if analysis_proppedcantilever == '1':
    print("You have selected Minimum Potential Energy Method.")

    print(
        "\n\n\n\nApproach:- We will get the true value of redundant reaction when the total strain energy in the beam will be minimum.")
    redundant = (3 * int(W) * int(L)) / 8
    print("\nRedundant reaction in the beam = R = (3*W*L)/8 = " + str(redundant) + " N")

    # User is asked whether he or she wishes to continue the analysis or not
    choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")

# If the method of analysis selected is Consistent Deformation Theorem Method then the following elif statement will be executed
elif analysis_proppedcantilever == '2':
    print("You have selected Consistent Deformation Theorem Method.")

    print("\n\n\n\nSTEP 01: CALCULATING DEGREE OF STATIC INDETERMINACY (Ds) ")
    print("\nNumber of actual reactions, R = 3")
    print("Number of equations available, E = 2")
    print("Degree of Static Indeterminacy, Ds = R - E = 1")

    print("\n\nSTEP 02: REMOVING REDUNDANT REACTIONS AND CALCULATING DISPLACEMENT DUE TO LOADING")
    print("\nLet 'Rb' be the redundant reaction at B.")
    deltaB1 = -(float(W) * float(L) * float(L) * float(L) * float(L)) / (8 * float(E) * float(I))
    print("Displacement at B due to applied loading, Delta(B1) = -(W*L^4)/(8*E*I) = " + str(
        deltaB1 * 0.001) + " m (Taking clockwise direction as negative)")

    print("\n\nSTEP 03: REMOVING LOADING AND CALCULATING DISPLACEMENT DUE TO REDUNDANT REACTIONS")
    Rb = (3 * float(W) * float(L)) / 8
    deltaB2 = (float(Rb) * float(L) * float(L) * float(L)) / (3 * float(E) * float(I))
    print("\nDisplacement at B due to redundant reaction at B, Delta(B2) = (Rb*L^3)/(3*E*I) = " + str(
        deltaB2 * 0.001) + " m")

    print("\n\nSTEP 04: OBTAINING COMPATIBILITY EQUATION AND SOLVING FOR 'Ma' and 'Mb'")
    print(
        "\nApproach:- To solve for redundant reaction, we need one more equation which can be obtained by compatability. It is quite evident that the displacement at B is zero.")
    print("Total displacement at B, Delta(B) = Delta(B1) + Delta(B2) = 0")
    print(
```

10 August 2023 06:53:53

```
            "Substituting the values and solving the above equation, we get redundant reaction at B, Rb = (3*W*L)/8 = " + str(
                Rb) + " N")
        Ra = (5 * float(W) * float(L)) / 8
        Ma = -(float(W) * float(L) * float(L)) / 8
        print("\nUsing the value of redundant reaction at B 'Rb', we can also calculate the following:")
        print("Reaction at A, 'Ra' = W*L - Rb = W*L - (3*W*L)/8 = (5*W*L)/8 = " + str(Ra) + " N")
        print("Bending Moment at A, 'Ma' = Rb - WL^2/2 = -(W*L^2)/8 = " + str(
            Ma * 0.001) + " N-m (Taking clockwise direction as negative)")

        # User is asked whether he or she wishes to continue the analysis or not
        choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")


# If the type of beam selected is simply supported beam then the following if statement will be executed
elif type_of_beam == '3':
    print("You have selected Simply Supported Beam.")

    # Loading conditions menu starts here
    print("\nThe various types of loading conditions available for Simply Supported beam are : ")

    # Point Load for Simply Supported beam
    print("\n1:-> Point Load :- ")
    print("               |P")
    print("               |")
    print("               V")
    print("       ----------------------")
    print("       ^       B        O")
    print("       A                C")
    print("\n")

    # Uniformly Distributed Load for Simply Supported beam
    print("2:-> Uniformly Distrbuted Load :- ")
    print("               |W")
    print("               |")
    print("               V")
    print("       /\/\/\/\/\/\/\/\/\/\/\/\/\")
    print("       ----------------------")
    print("       ^                O")
```

10 August 2023 06:53:53

```
print("       A                B")
print("\n")

# User is asked to select the type of loading condition he or she wishes to apply to the beam
load_simplysupported = input("Please specify the loading condition (1 or 2) : ")

# If the type of loading condition selected is point load then the following if statement will be executed
if load_simplysupported == '1':
    print("You have selected Point Load as the loading condition.")

    # User is asked to enter values for load, length of beam, modulus of elasticity and moment of inertia
    P = input("\nEnter load magnitude (P) in N : ")
    L = input("Enter length of beam (L) in mm : ")
    E = input("Enter modulus of elasticity (E) of the material of beam in N/mm^2 : ")
    I = input("Enter moment of inertia (I) of beam in mm^4 : ")

    # Menu for various structural analysis methods starts here
    print("\nThe various methods of analysis available are : ")
    print("1:-> Strain Energy Method")

    # User is asked to select the method of structural analysis
    analysis_simplysupported = input("\nPlease select the method of analysis : ")

    # If the method of analysis selected is Strain Energy Method then the following if statement will be executed
    if analysis_simplysupported == '1':
        print("You have selected Strain Energy Method.")

        print("\n\n\n\nSTEP 01 : CALCULATING STRAIN ENERGY (u)")
        u = (float(P) * float(P) * float(L) * float(L) * float(L)) / (96 * float(E) * float(I))
        print("\nStrain energy in this case, represented by u, is given as = (P^2*L^3)/(96*E*I) = " + str(
            u * 0.001) + " N-m")

        print("\n\nSTEP 02 : FINDING DEFLECTION AT POINT 'B'")
        deflection = (float(P) * float(L) * float(L) * float(L)) / (48 * float(E) * float(I))
        print("\nDeflection at point 'B' is given as = (P*L^3)/(48*E*I) = " + str(deflection * 0.001) + " m")

        # User is asked whether he or she wishes to continue the analysis or not
        choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")
```

10 August 2023 06:53:53

```
    # If the type of loading condition selected is uniformly distributed load then the following elif statement will be executed
    else:
        print("You have selected Uniformly Distributed Load as the loading condition.")

        # User is asked to enter values for uniformly distributed load, length of beam, modulus of elasticity and moment of inertia
        W = input("\nEnter uniformly distributed load (w) in N/mm : ")
        L = input("Enter length of beam (L) in mm : ")
        E = input("Enter modulus of elasticity (E) of the material of beam in N/mm^2 : ")
        I = input("Enter moment of inertia (I) of beam in mm^4 : ")

        # Menu for various structural analysis methods starts here
        print("\nThe various methods of analysis available are : ")
        print("1:-> Strain Energy Method")

        # User is asked to select the method of structural analysis
        analysis_simplysupported = input("\nPlease select the method of analysis : ")

        # If the method of analysis selected is Strain Energy Method then the following if statement will be executed
        if analysis_simplysupported == '1':
            print("You have selected Strain Energy Method.")

            print("\n\n\n\nSTEP 01 : CALCULATING STRAIN ENERGY (u)")
            u = (float(W) * float(W) * float(L) * float(L) * float(L) * float(L) * float(L)) / (
                    96 * float(E) * float(I))
            print("\nStrain energy in this case, represented by u, is given as = (W^2*L^5)/(96*E*I) = " + str(
                u * 0.001) + " N-m")

            print("\n\nSTEP 02 : FINDING DEFLECTION AT POINT 'B'")
            deflection = (float(W) * float(L) * float(L) * float(L) * float(L)) / (48 * float(E) * float(I))
            print("\nDeflection at point 'B' is given as = (W*L^$)/(48*E*I) = " + str(deflection * 0.001) + " m")

            # User is asked whether he or she wishes to continue the analysis or not
            choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")

# If the type of beam selected is fixed beam then the following else statement will be executed
else:
    print("You have selected Fixed Beam.")
```

10 August 2023 06:53:53

```
# Loading conditions menu starts here
print("\nThe various types of loading conditions available for fixed beam are : ")

# Point Load for fixed beam
print("\n1:-> Point Load :- ")
print("                |P")
print("                |")
print("                V")
print("       |----------------------|")
print("       A                      B")
print("\n")

# Uniformly Distributed Load for cantilever beam
print("2:-> Uniformly Distributed Load :- ")
print("                |W")
print("                |")
print("                V")
print("        /\/\/\/\/\/\/\/\/\/\/\/\")
print("       |----------------------|")
print("       A                      B")
print("\n")

# User is asked to select the type of loading condition he or she wishes to apply to the beam
load_fixed = input("Please specify the loading condition (1 or 2) : ")

# If the type of loading condition selected is point load then the following if statement will be executed
if load_fixed == '1':
    print("You have selected Point Load as the loading condition.")

    # User is asked to enter values for load, length of beam, modulus of elasticity and moment of inertia
    P = input("\nEnter load magnitude (P) in N : ")
    L = input("Enter length of beam (L) in mm : ")
    E = input("Enter modulus of elasticity (E) of the material of beam in N/mm^2 : ")
    I = input("Enter moment of inertia (I) of beam in mm^4 : ")

    # Menu for various structural analysis methods starts here
    print("\nThe various methods of analysis available are : ")
```

10 August 2023 06:53:53

```
print("1:-> Consistent Deformation Method")

# User is asked to select the method of structural analysis
analysis_fixed = input("\nPlease select the method of analysis : ")

# If the method of analysis selected is Consistent Deformation Method then the following elif statement will be executed
if analysis_fixed == '1':
    print("You have selected Consistent Deformation Theorem Method.")

    print("\n\n\n\nSTEP 01: CALCULATING DEGREE OF STATIC INDETERMINACY (Ds) ")
    print("\nNumber of actual reactions, R = 4")
    print("Number of equations available, E = 2")
    print("Degree of Static Indeterminacy, Ds = R - E = 2")

    print("\n\nSTEP 02: REMOVING REDUNDANT REACTIONS AND CALCULATING DISPLACEMENTS DUE TO LOADING")
    print("\nLet 'Ma' be the redundant reaction at A and 'Mb' be the redundant reaction at B.")
    thetaA1 = (float(P) * float(L) * float(L)) / (16 * float(E) * float(I))
    thetaB1 = -(float(P) * float(L) * float(L)) / (16 * float(E) * float(I))
    print("Displacement at A due to applied loading, Theta(A1) = (P*L^2)/(16*E*I) = " + str(thetaA1))
    print("Displacement at B due to applied loading, Theta(B1) = -(P*L^2)/(16*E*I) = " + str(
        thetaB1) + " (Taking clockwise direction as negative)")

    print("\n\nSTEP 03: REMOVING LOADING AND CALCULATING DISPLACEMENTS DUE TO REDUNDANT REACTIONS")
    print("\nLet 'Ma' be the redundant reaction at A and 'Mb' be the redundant reaction at B.")
    Ma = (float(P) * float(L)) / 8
    thetaA2 = -(float(Ma) * float(L)) / (3 * float(E) * float(I))
    thetaB2 = (float(Ma) * float(L)) / (6 * float(E) * float(I))
    print("\nDisplacement at A due to redundant reaction at A, Theta(A2) = -(Ma*L)/(3*E*I) = " + str(
        thetaA2) + " (Taking clockwise direction as negative)")
    print("Displacement at B due to redundant reaction at A, Theta(B2) = (Ma*L)/(6*E*I) = " + str(thetaB2))
    Mb = (float(P) * float(L)) / 8
    thetaA3 = -(float(Mb) * float(L)) / (6 * float(E) * float(I))
    thetaB3 = (float(Mb) * float(L)) / (3 * float(E) * float(I))
    print("Displacement at A due to redundant reaction at B, Theta(A3) = -(Mb*L)/(6*E*I) = " + str(
        thetaA3) + " (Taking clockwise direction as negative)")
    print("Displacement at B due to redundant reaction at B, Theta(B3) = (Mb*L)/(3*E*I) = " + str(thetaB3))

    print("\n\nSTEP 04: OBTAINING COMPATIBILITY EQUATION AND SOLVING FOR 'Ma' and 'Mb'")
```

10 August 2023 06:53:53

```
        print(
            "\nApproach:- To solve for redundant reactions, we need one more equation which can be obtained by compatability. It is quite evident that the displacement at the fixed supports A and B is zero.")
        print("Therefore, total displacement at A, Theta(A) = Theta(A1) + Theta(A2) + Theta(A3) = 0")
        print("Also, total displacement at B, Theta(B) = Theta(B1) + Theta(B2) + Theta(B3) = 0")
        print(
            "\nSolving the above two equations, we get redundant reaction at A, Ma = (P*L)/8 = " + str(Ma*0.001) + " N-m")
        print("Redundant reaction at B, Mb = (P*L)/8 = " + str(Mb*0.001) + " N-m")

        # User is asked whether he or she wishes to continue the analysis or not
        choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")


# If the type of loading condition selected is uniformly distributed load then the following else statement will be executed
else:
    print("You have selected Uniformly Distributed Load as the loading condition.")

    # User is asked to enter values for uniformly distributed load, length of beam, modulus of elasticity and moment of inertia
    W = input("\nEnter uniformly distributed load (w) in N/mm : ")
    L = input("Enter length of beam (L) in mm : ")
    E = input("Enter modulus of elasticity (E) of the material of beam in N/mm^2 : ")
    I = input("Enter moment of inertia (I) of beam in mm^4 : ")

    # Menu for various structural analysis methods starts here
    print("\nThe various methods of analysis available are : ")
    print("1:-> Consistent Deformation Theorem Method")

    # User is asked to select the method of structural analysis
    analysis_fixed = input("\nPlease select the method of analysis : ")

    # If the method of analysis selected is Consistent Deformation Theorem Method then the following elif statement will be executed
    if analysis_fixed == '1':
        print("You have selected Consistent Deformation Theorem Method.")

        print("\n\n\n\nSTEP 01: CALCULATING DEGREE OF STATIC INDETERMINACY (Ds) ")
        print("\nNumber of actual reactions, R = 4")
        print("Number oe equations available, E = 2")
        print("Degree of Static Indeterminacy, Ds = R - E = 2")
```

10 August 2023 06:53:53

```
        print("\n\nSTEP 02: REMOVING REDUNDANT REACTIONS AND CALCULATING DISPLACEMENTS DUE TO LOADING")
        print("\nLet 'Ma' be the redundant reaction at A and 'Mb' be the redundant reaction at B.")
        thetaA1 = (float(W) * float(L) * float(L) * float(L)) / (24 * float(E) * float(I))
        thetaB1 = -(float(W) * float(L) * float(L) * float(L)) / (24 * float(E) * float(I))
        print("Displacement at A due to applied loading, Slope(A1) = (W*L^3)/(24*E*I) = " + str(thetaA1))
        print("Displacement at B due to applied loading, Slope(B1) = -(W*L^3)/(24*E*I) = " + str(
            thetaB1) + " (Taking clockwise direction as negative)")

        print("\n\nSTEP 03: REMOVING LOADING AND CALCULATING DISPLACEMENT DUE TO REDUNDANT REACTIONS")
        print("\nLet 'Ma' be the redundant reaction at A and 'Mb' be the redundant reaction at B.")
        Ma = (float(W) * float(L) * float(L)) / 12
        thetaA2 = -(float(Ma) * float(L)) / (3 * float(E) * float(I))
        thetaB2 = (float(Ma) * float(L)) / (6 * float(E) * float(I))
        print("\nDisplacement at A due to redundant reaction at A, Slope(A2) = -(Ma*L)/(3*E*I) = " + str(
            thetaA2) + " (Taking clockwise direction as negative)")
        print("Displacement at B due to redundant reaction at A, Slope(B2) = (Ma*L)/(6*E*I) = " + str(thetaB2))
        Mb = (float(W) * float(L) * float(L)) / 12
        thetaA3 = -(float(Mb) * float(L)) / (6 * float(E) * float(I))
        thetaB3 = (float(Mb) * float(L)) / (3 * float(E) * float(I))
        print("Displacement at A due to redundant reaction at B, Slope(A3) = -(Mb*L)/(6*E*I) = " + str(
            thetaA3) + " (Taking clockwise direction as negative)")
        print("Displacement at B due to redundant reaction at B, Slope(B3) = (Mb*L)/(3*E*I) = " + str(thetaB3))

        print("\n\nSTEP 04: OBTAINING COMPATIBILITY EQUATION AND SOLVING FOR 'Ma' and 'Mb'")
        print(
            "\nApproach:- To solve for redundant reactions, we need one more equation which can be obtained by compatability. It is quite
evident that the displacement at the fixed supports A and B is zero.")
        print("Therefore, total displacement at A, Slope(A) = Slope(A1) + Slope(A2) + Slope(A3) = 0")
        print("Also, total displacement at B, Slope(B) = Slope(B1) + Slope(B2) + Slope(B3) = 0")
        print("\nSolving the above two equations, we get redundant reaction at A, Ma = (W*L^2)/12 = " + str(
            Ma*0.001) + " N"-m)
        print("Redundant reaction at B, Mb = (W*L^2)/12 = " + str(Mb*0.001) + " N-m")

        # User is asked whether he or she wishes to continue the analysis or not
        choice = input("\n\n\n\nPlease enter 'y' if you wish to continue and 'n' to discontinue the anlaysis : ")

# If the user doesn't wish to continue with the program for beam analysis then the following else statement will be executed
```

10 August 2023 06:53:53

```
else:
    print(exit)
    exit()
```

RESULTS AND DISCUSSION

Some of the above cases are verified as below. It is done by comparing manual calculations with the python model. Here, it is considered that the user first performs case 1b of Cantilever beam with UDL using Strain Energy Method. Next, before closing the program, the user also performs Case 4a. The input given by the user are as below.

Input

For case 1b – cantilever beam – UDL using strain energy method

User input
load magnitude (w) in N/mm : 1
length of beam (L) in mm : 5000
modulus of elasticity (E) of the material of beam in N/mm^2 : 200000
moment of inertia (I) of beam in mm^4 : 81960000

For case 4a – fixed beam – point load using consistent deformation method

User Input
load magnitude (P) in N : 100
length of beam (L) in mm : 20000
modulus of elasticity (E) of the material of beam in N/mm^2 : 199947.96
moment of inertia (I) of beam in mm^4 : 83246286

Output

On compiling and running the Python script or code, a part of the total output is represented below corresponding to the cases as inputted above. It must be noted that output screen will display the sub classification of beam conditions (sub – case a and b) in the form of diagram after selection of type of beam (case 1, 2, 3, 4) from which the user can make a choice. After which, the values for load, length of beam, modulus of elasticity, and moment of inertia will be asked to input. Once, the case analysis is completed, the user is given the option to either continue analyzing for same or different cases or exit the program from there. In the context of inputs given in the above section, the user will first analyze for case 1b and then continue for case 4a before exiting the program.

Case 1b

```
Enter uniformly distributed load (w) in N/mm : 1
Enter length of beam (L) in mm : 5000
Enter modulus of elasticity (E) of the material of beam in N/mm^2 : 200000
Enter moment of inertia (I) of beam in mm^4 : 81960000

The various methods of analysis available are :
1:-> Strain Energy Method

Please select the method of analysis : 1
You have selected Strain Energy Method.

STEP 01 : CALCULATING STRAIN ENERGY (u)

Strain energy in this case, represented by u, is given as = (W^2*L^4)/(8*E*I) = 0.004766044411908248 N-m


STEP 02 : FINDING DEFLECTION AT POINT 'B'

Deflection at point 'B' is given as = (5*W*L^4)/(48*E*I) = 0.003971703676590207 m
```

10 August 2023 06:53:53

Case 4a

Enter load magnitude (P) in N : 100
Enter length of beam (L) in mm : 20000
Enter modulus of elasticity (E) of the material of beam in N/mm^2 : 199947.96
Enter moment of inertia (I) of beam in mm^4 : 83246286

The various methods of analysis available are :
1:-> Consistent Deformation Method

Please select the method of analysis : 1
You have selected Consistent Deformation Theorem Method.

STEP 01: CALCULATING DEGREE OF STATIC INDETERMINACY (Ds)

Number of actual reactions, R = 4
Number of equations available, E = 2
Degree of Static Indeterminacy, Ds = R - E = 2

STEP 02: REMOVING REDUNDANT REACTIONS AND CALCULATING DISPLACEMENTS DUE TO LOADING

Let 'Ma' be the redundant reaction at A and 'Mb' be the redundant reaction at B.
Displacement at A due to applied loading, Theta(A1) = (P*L^2)/(16*E*I) = 0.0001501959300204788
Displacement at B due to applied loading, Theta(B1) = -(P*L^2)/(16*E*I) = -0.0001501959300204788 (Taking clockwise direction as negative)

STEP 03: REMOVING LOADING AND CALCULATING DISPLACEMENTS DUE TO REDUNDANT REACTIONS

Let 'Ma' be the redundant reaction at A and 'Mb' be the redundant reaction at B.

Displacement at A due to redundant reaction at A, Theta(A2) = -(Ma*L)/(3*E*I) = -0.00010013062001365254 (Taking clockwise direction as negative)
Displacement at B due to redundant reaction at A, Theta(B2) = (Ma*L)/(6*E*I) = 5.006531000682627e-05
Displacement at A due to redundant reaction at B, Theta(A3) = -(Mb*L)/(6*E*I) = -5.006531000682627e-05 (Taking clockwise direction as negative)
Displacement at B due to redundant reaction at B, Theta(B3) = (Mb*L)/(3*E*I) = 0.00010013062001365254

STEP 04: OBTAINING COMPATIBILITY EQUATION AND SOLVING FOR 'Ma' and 'Mb'

Approach:- To solve for redundant reactions, we need one more equation which can be obtained by compatability. It is quite evident that the displacement at the fixed supports A and B is zero.
Therefore, total displacement at A, Theta(A) = Theta(A1) + Theta(A2) + Theta(A3) = 0
Also, total displacement at B, Theta(B) = Theta(B1) + Theta(B2) + Theta(B3) = 0

Solving the above two equations, we get redundant reaction at A, Ma = (P*L)/8 = 250.0 N-m
Redundant reaction at B, Mb = (P*L)/8 = 250.0 N-m

Verification

Comparison of output with manual calculation using Table 1 for both cases were performed which are as discussed below.

For case 1b

Output got from the Python model on successful analysis of case 1.b is as below.

1. Strain Energy = 0.004766044411908248 N-m
2. Deflection at B = 0.003971703676590207 m

Manual Calculation (Using Table 1)
U = 0.0048 N-m
δ_B = 0.004 m

10 August 2023 06:53:53

Since, final result from program matches with the manual calculations as given in Table 1.
Hence, verified.
It is also observed that the result got from the Python model are much more accurate.

For case 4a

From the above code, the analysis results got were as below.

1. Moment at A = 250.0 N-m
2. Moment at B = 250.0 N-m

Manual Calculation
M_A = 250.00 N-m
M_B = 250.00 N-m

Since, final result from program matches with the calculations.
Hence, verified.

CONCLUSION

The following conclusions may be drawn from this paper:

- Proposed a significant practical model for static analysis of cantilever, propped cantilever, simply supported, and fixed beams having point loads and Uniformly Distributed load (UDL) for the inputted values of Young's modulus of elasticity, the moment of inertia, and loads.
- It was found that the outputs got from the proposed code were in good agreement with the manual calculations as discussed in the Verification section.
- Through this paper, an attempt has been made to foster the scope of structural engineering in the programming language - Python.
- Computer languages are all about implementing algorithms. And by breaking down the analysis of beams into multiple cases and laying the procedure for each case further in a stepwise approach, the analysis was formulated into an algorithm that made possible a task that is more organized, fast, and accurate by avoiding any human error.

10 August 2023 06:53:53

FUTURE SCOPE OF THE PAPER

The future scope of the paper is to apply these computational techniques to large sets of real measurements obtained via industry partner(s) or relating to university buildings. In future, this paper will be taken forward and will be made inclusive of more beam types and diverse loading conditions. We also envision to venture into analyzing different structural members in coming times.

Data Availability Statement

All data, models and code generated or used during the study appear in the submitted paper.

Funding Statement

Authors would like to express their sincere gratitude to Delhi Technological University, Delhi, India for providing funding to conduct the research work.

ACKNOWLEDGEMENT

Authors would like to convey the exceptional thanks and appreciation to our Institute DTU, Delhi for providing resources, and **Dr. Manvendra Verma** encouraged the authors to submit the paper within the allowable time and provided his valuable comments to finalize this paper.

REFERENCES

1. N. Gaur, R. Raj, and P. K. Goyal, "Interference effect on corner - configured structures with variable geometry and blockage configurations under wind loads using CFD," *Asian Journal of Civil Engineering*, no. October, 2021, doi: 10.1007/s42107-021-00400-0.

2. A. Kumar and R. Raj, "Study of Pressure Distribution on an Irregular Octagonal Plan Oval-Shape Building Using CFD," *Civil Engineering Journal*, vol. 7, no. 10, pp. 1787–1805, Oct. 2021, doi: 10.28991/cej-2021-03091760.

3. N. Gaur and R. Raj, "Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel," *Ain Shams Engineering Journal*, no. xxxx, 2021, doi: 10.1016/j.asej.2021.06.007.

4. R. K. Meena, R. Raj, and S. Anbukumar, *Numerical Investigation of Wind Load on Side Ratio of High-Rise Buildings Numerical Investigation of Wind Load on Side Ratio of High-Rise Buildings*, no. December. Springer Singapore, 2021.

5. R. K. MEENA, G. P. Awadhiya, A. P. Paswan, and H. K. Jayant, "Effects of Bracing System on Multistoryed Steel Building," 2018, doi: 10.1088/1757-899X/1128/1/012017.

6. R. Raj and A. K. Ahuja, "Wind Loads on Cross Shape Tall Buildings," *Journal of Academia and Industrial Research (JAIR)*, vol. 2, no. 2, pp. 111–113, 2013.

7. S. Mahajan, V. Yadav, R. Raj, and R. Raj, *Effect of Shear Walls on Tall Buildings with Different Corner Configuration Subjected to Wind Loads*. Springer Singapore.

8. R. Raj, A. Sharma, and S. Chauhan, "Response of Square and Plus Shaped Buildings on Varying Wind Loads," pp. 206–215, 2018, doi: 10.1061/9780784482032.022.

9. C. A. Shaffer and V. Tech, "Data Structures and Algorithm Analysis," vol. 2, 2011.

10. A. Neubauer, J. Freudenberger, and V. Kühn, "Coding Theory: Algorithms, Architectures, and Applications," Coding Theory: Algorithms, Architectures, and Applications. John Wiley & Sons Ltd, pp. 1–340, 2007, doi: 10.1002/9780470519837.

11. *Chapter 1 Pointers , Arrays , and Structures*. 1996.

12. Z. A. Shaw, *Learn PYTHON 3 the Hard Way*. Addison Wesley, 1375.

13. R. Johansson, *Numerical python: Scientific computing and data science applications with numpy, SciPy and matplotlib, Second edition*. 2018.

14. I. A. Karnovsky, *Theory of Arched Structures*. 2012.

15. J. J. Connor and S. Faraji, *Fundamentals of Structural Engineering*. 2013.

16. R. C. Hibbeler, *Structural Analysis*, 8th ed. Peasrson, 2012.

17. B. G. Nielson and J. C. Mccormac, *Structural Analysis*. United state of America: John Wiley & Sons Ltd, 2017.

18. K. Derucher, C. Putcha, and U. Kim, *Indeterminate structural analysis*, vol. 1, no. 1. 2013.

10 August 2023 06:53:53

19. B. Koo Lee, J. F. Wilson, and S. Jin Oh, “Elastica of cantilevered beams with variable cross sections,” *International Journal of Non-Linear Mechanics*, vol. 28, no. 5, pp. 579–589, 1993, doi: 10.1016/0020-7462(93)90049-Q.

20. M. C. Ece, M. Aydogdu, and V. Taskin, “Vibration of a variable cross-section beam,” *Mechanics Research Communications*, vol. 34, no. 1, pp. 78–84, 2007, doi: 10.1016/j.mechrescom.2006.06.005.

21. E. J. Sapountzakis and G. C. Tsiatas, “Elastic flexural buckling analysis of composite beams of variable cross-section by BEM,” *Engineering Structures*, vol. 29, no. 5, pp. 675–681, 2007, doi: 10.1016/j.engstruct.2006.06.010.

22. E. J. Sapountzakis and G. C. Tsiatas, “Flexural buckling analysis of composite beams of variable cross-section by BEM,” *International Conference on Computational Methods for Copupled Problmes in Science and Engineering*, vol. 29, no. 5, pp. 675–681, 2005, doi: 10.1016/j.engstruct.2006.06.010.

23. M. A. Rahman and M. A. Kowser, “Nonlinear analysis of cantilever shape memory alloy beams of variable cross section,” *Smart Materials and Structures*, vol. 16, no. 2, pp. 531–540, 2007, doi: 10.1088/0964-1726/16/2/035.

24. M. Boiangiu, V. Ceausu, and C. D. Untaroiu, “A transfer matrix method for free vibration analysis of Euler-Bernoulli beams with variable cross section,” *JVC/Journal of Vibration and Control*, vol. 22, no. 11, pp. 2591–2602, 2014, doi: 10.1177/1077546314550699.

25. A. L. Rojas and J. M. Ramirez, “Mathematical model for rectangular beams of variable cross section of symmetrical linear shape for uniformly distributed load,” vol. 10, no. 2, pp. 545–564, 2014.

26. P. Dumitrache, “Shear Stresses in Beams with Variable Cross Section, Subjected to Non-Uniform Bending,” pp. 39–44, 2012.

27. E. J. Sapountzakis and D. G. Panagos, “Nonlinear analysis of beams of variable cross section, including shear deformation effect,” *Archive of Applied Mechanics*, vol. 78, no. 9, pp. 687–710, 2008, doi: 10.1007/s00419-007-0182-5.

10 August 2023 06:53:53

2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) | 979-8-3503-1590-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACCAI58221.2023.10199737

Exploring the Efficacy of Deep Learning Models for Multiclass Toxic Comment Classification in Social Media Using Natural Language Processing

Prashant Giridhar Shambharkar, Hastej Singh, Harshit Raj Raghav and Hilansh Verma

Department of Computer Engineering Delhi Technological University Delhi, India

E-mail : prashant.shambharkar@dtu.ac.in hastejsingh694@gmail.com harshitrajraghav@gmail.com hilansh001@gmail.com

Abstract- Our thesis focuses on developing a deep learning-based toxic comment classifier. The classifier will be used to find and report potentially unpleasant or hazardous information on websites like social media, discussion boards, and comment sections. We want to encourage a welcoming and safe online community and stop the growth of online abuse such as cyberbullying and hate speech. Long Short-Term Memory (LSTM) and Hybrid LSTM-CNN (Convolutional Neural Network and LSTM based Approach) are two algorithms that the classifier uses to categorize the comments depending on their level of toxicity, such as threats, obscenity, insults, and identity-based hatred. The classifier's input data came from Kaggle and underwent a number of pre-processing processes, including lemmatization and normalizing the text data.

Keywords- CNN, LSTM, LSTM-CNN, NLP, Toxic Comment.

I.INTRODUCTION

The rise of social media and online communication has provided an unprecedented level of connectivity for individuals around the world. However, with this increase in communication comes the challenge of dealing with the negative aspects of online discourse, including cyberbullying, hate speech, and other forms of toxic comments. Toxic comment categorization, the process of identifying and highlighting potentially harmful or objectionable content, has become an increasingly important tool for online platforms to promote a secure and welcoming environment for users.

Toxic comments can take many forms, including explicit language, threats, insults, and hate speech targeted at specific individuals or groups. The propagation of these types of comments can create an environment that is challenging for people to express themselves or engage in healthy dialogue online. Furthermore, online platforms have a responsibility to protect their users from abusive behavior and harassment and may even face liability for failing to take action.

To address these challenges, we present a deep learning-based approach for toxic comment classification. Our approach employs two distinct techniques, LSTM and LSTM-CNN, to identify and categorize toxic comments according to their type of toxicity. Specifically, we focus on four categories of toxicity: toxic, severe toxic, obscene, threats, insults, and identity hate speech based on an individual's identification.

The goal of this research is to contribute to the development of effective toxic comment categorization systems that can be used by online platforms to promote a safer and more welcoming online environment. Our approach builds on existing research in deep learning and natural language processing and provides a framework for the development of more sophisticated and accurate toxic comment classifiers.

The rest of the paper arranged as follows: Section 2 discuss Related work; Section 3 preliminary related to the algorithms employed in the dataset; Section 4 presents the methodology; Section 5 concludes the paper.

II.RELATED WORK

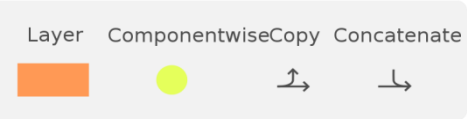
Classifying toxic comments is a significant problem belonging to natural language processing (NLP) realm that has got much recognition in past few years. Various types of Machine learning models such as convolutional neural networks (CNNs) and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:25:02 UTC from IEEE Xplore. Restrictions apply.

recurrent neural networks (RNNs), have given promising outcome in this area. Here, we review some of the most relevant works in the field.

One of the earliest works on using RNNs for toxic comment classification was by Zhang et al. (2018)[1], who proposed a bi-directional LSTM model with attention mechanisms. Their model was able to give promising performance on the Wikipedia Toxicity dataset. A similar approach was taken by Chen et al. (2019)[2], who proposed a bi-directional RNN with a self-attention mechanism.

More recently, several works have proposed hybrid models that combine RNNs and CNNs to leverage the strengths of both architectures. Another approach to toxic comment classification involves the use of transfer learning. Devlin et al. (2019)[3] proposed the BERT model, which uses a transformer architecture to perform contextualized word embeddings. The authors showed that their model achieved the current highest level of performance or technology in a particular field or domain. Similarly, Sun et al. (2020)[4] fine-tuned BERT on a large-scale dataset of toxic comments and achieved state-of-the-art performance on several benchmark datasets.



Other researchers have explored various approaches for toxic comment classification. Qian et al. (2020)[5] proposed a hierarchical model that uses both character-level and word-level embedding to classify toxic comments. The authors demonstrated that their model outperformed existing models on multiple standard datasets, establishing a new current highest level of performance or technology in a particular field or domain. Zhang et al. (2021)[6] proposed a multi-task learning approach that jointly learns to classify toxic comments and identify the users who posted them.

Several works have also focused on exploring different aspects of the problem, like identifying offensive language and expression of hatred. Davidson et al. (2017) [7] proposed a dataset and a set of classification models to detect hate speech on social media. Waseem and Hovy (2016) [8] proposed a dataset and a set of classification models to detect offensive language on Twitter.

In summary, the field of toxic comment classification using deep learning models is quickly evolving, with new emerging techniques and datasets being continuously developed and tested all the time. The works reviewed here represent some of the most relevant and promising approaches in the field.

III. PRELIMINARIES

The models that we have used in our paper are listed as following:

A. LTSM

LSTM, or Long Short-Term Memory, is a part of the Recurrent Neural Networks (RNNs), and is particularly adept at identifying long-term dependencies within sequential data. It was created to get around the difficulties regular RNNs have with memorizing data over lengthy sequences.

An LSTM network features a memory cell that can store data and regulate the movement of data across the network. Gates in the network also control how information enters and exits memory cells. As a result, LSTMs can remember important information while forgetting less critical information.

Natural language processing, speech recognition, and even stock market prediction are just a few of the uses for LSTMs. For processing sequential data, where the sequence of the data points is essential, they have had great success. Also, LSTMs are perfect for situations where both forms of information are crucial since they can handle both short-term and long-term dependencies.

Overall, LSTMs have been widely used in numerous industries and have shown to be a successful method for processing of sequential data.

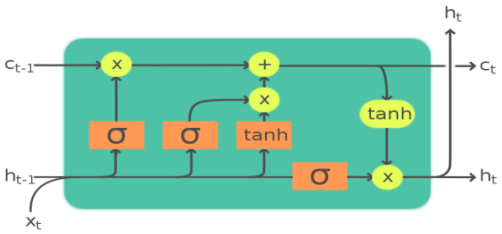


Fig. 1 Long Short Term Memory[14]

B. LSTM-CNN

Two popular Deep Learning Architectures Long

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:25:02 UTC from IEEE Xplore. Restrictions apply.

Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) are combined to form LSTM-CNN (CNN). This model provides the benefits of CNNs and LSTMs to solve certain problems with sequence data.

While processing sequential data, LSTMs perform well since the sequence of the data points matters. Long-term dependencies may be remembered by them, but they have trouble digesting a lot of information.

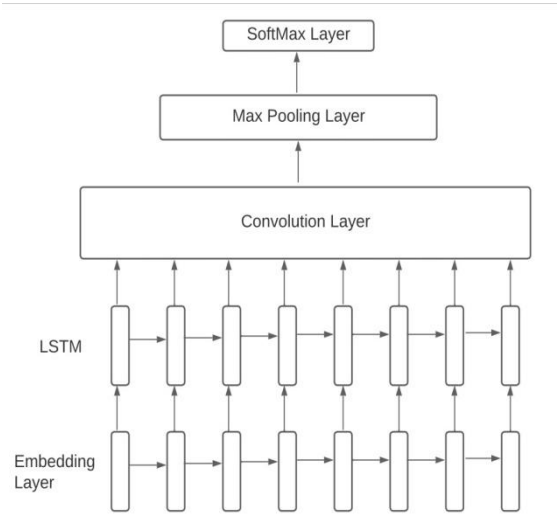


Fig. 2 LSTM-CNN Hybrid Model

On the other hand, CNNs can extract features from unprocessed input and are built to analyze enormous volumes of data. They have trouble with sequences, though, because they don't take the order of the data points into account[15].

Combining these advantages, the LSTM-CNN model processes sequential data with LSTMs and using CNNs for feature extraction from unprocessed data. From the input data, the CNNs extract useful characteristics and send them to the LSTM layers for processing. To create predictions, the LSTM layers combine the sequential data with the features from the CNNs.

For applications including sentiment analysis, picture captioning, and language translation, LSTM-CNN models are used. They have shown to be efficient in managing enormous volumes of data while taking the order of the data points into account.

IV. Proposed Methodology and Result Discussion

The training and test sets were included in the data that was taken from Kaggle and utilized to build the categorization model. The entire process is outlined in the workflow that follows, which goes from Model Assessment through Data Pre-Processing.

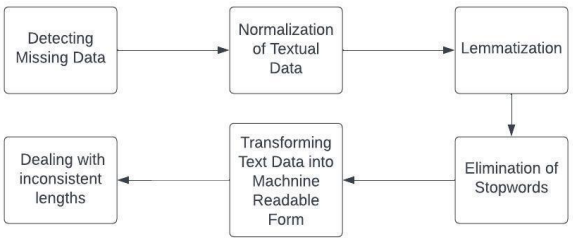


Fig. 3 Data pre-processing and exploration flowchart

C. Data Pre-processing and Exploration

1. *Step 1: Detecting Missing Data.*

The training and test data were initially imported into a pandas dataframe. We needed to make sure there were no missing numbers in the data before moving on. There were no records with missing data when we used the "isnull" function on the set of data that is used for training and testing reasons. Afterward, we continued with the next stage of our project.

2. *Step 2: Normalization of Textual Data.*

We started with the data pre-processing stage after making sure the data was clear of any missing values. As comments from internet forums frequently contain inconsistent grammar, unusual characters, and digits that represent letters, the first step was to normalize the text data. We used Regex to correct these discrepancies.

We used jupyter notebooks to carry out these tasks. We started by compiling a list of popular swear words that are often used on social networking sites or online forums. Then, we created a function that executes all of the previous procedures and produces accurate data. Both the training and test datasets underwent these stages.

The following is a list of the text normalization steps we carried out:

TABLE 1 STEPS FOR NORMALIZATION

Normalization Steps:
a) Deleting characters within text.
b) Deleting duplicate characters.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:25:02 UTC from IEEE Xplore. Restrictions apply.

c) Conversion of all the data to lowercase.
d) Dropping punctuation symbols.
e) Removing extra white spaces between words.
f) Eliminating the "\n" character.
g) Deleting non-English characters.

3. *Step 3: Lemmatization.*

Lemmatization procedure is started after the cleaning and normalizing the text data. Lemmatization is combining the various variations of a word so that the machine learning algorithm can consider them as a single object. This decreases the dimensionality of the data and lowers the count of distinct words in the data. We used the "WordNetLemmatizer" from the "nltk" library to do lemmatization and made the function "lemma" to carry out the operation. To get lemmatized data, this function was used to the cleaned data acquired in Step 2.

4. *Step 4: Elimination of Stopwords.*

Eliminating stopwords is a critical step in text classification tasks because it concentrates attention on words that determine the text's meaning

(a) The "spacy" package was used to remove stopwords from our data. A list of typical stopwords included in the library named "STOP WORDS" may be used to eliminate stopwords from any text.

(b) Despite the spacy library's comprehensive list,We chose to search for stopwords that were specific to our dataset.

(c) To start, we expanded the list of stopwords by including single- and double-letter words. We discovered instances of single-letter or two-letter words in our dataset when evaluating random remarks, such as Wow what a gorgeous cushion, or He is such a cheerful man, bb. We included these single-letter or two-letter words in our list of stopwords.to make sure they don't affect how well our deep learning model performs. However, we made sure to leave out from the list of stopwords letters like "I" and "a" as well as words like "me," "am," and "as."

(d) Following the completion of this work, we looked for terms in our dataset that may be stopwords (based on their high frequency in the dataset and limited contribution to the classification task). To do this, we wrote a code snippet that enabled us to locate stopwords in our dataset that met these requirements. In order to generate our final stopwords list, we finally added these newly found stopwords to the spacy "STOP WORDS" list.

(e) We removed stopwords from both our training and test sets of data using our final stopwords list, producing a clean dataset free of errors.

5. *Step 5: Transforming Text Data into Machine Readable Form*

It is widely known that, regardless of the nature of the problem, Models utilizing both conventional machine learning techniques and deep learning methodologies handle numerical data. And therefore, in order to train and create a deep learning model which utilizes text inputs that are free of any errors or inconsistencies, it needs to be converted into a language that computers can understand. To do this, the following actions are required:

Tokenization: The text is broken up into separate words, thus "I love cats and love dogs" becomes ["We", "like", "lions", "and," "tigers"].

Indexing - Words are arranged in a manner similar to a dictionary and given an index, such as {1: "We", 2: "like", 3: "lions", 4: "and", 5: "tigers"}.

Index Representation - The order of the text's words may be represented as an index, which then can be employed to train a Deep Learning model. Examples [1, 2, 3, 4, 2, 5].

The "Tokenizer" class from the "Keras" library makes it easy to perform these actions. This technique can convert each text into a set of integers, where each different integer corresponds to the index of a token present in a dictionary, or into a vector where each token can be represented by a binary coefficient based on tf-idf or word count. This class is useful for converting text data into a vectorized format.

6. *Step 6: Dealing with inconsistent lengths*

We address the issue of inconsistent lengths of sequences in our text data by using padding. The lengths of comments Statements that are present on social media or online discussion platforms can vary greatly, from one-word responses to detailed elaborations. These varying length sentences are transformed into sequences of varying length vectors and cannot be input into deep learning models. To resolve this issue, padding is used to make the shorter sequences equal in length to the longer ones. In this case, the sentence length was established at 200 words with post-padding by utilizing the "pad_sequences" function from the Keras library. The padding procedure included adding zeros to the end of shorter sequence vectors. The subsequent step involves creating deep learning models.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:25:02 UTC from IEEE Xplore. Restrictions apply.

D. Construction and Evaluation of the Models

1. *Step 1: Segregate the Data into Training and Testing subsets.*

Once the feature Engineering and data pre-processing tasks are completed, we can move on to constructing and assessing our model. Before training a deep learning model on the training data, it is advisable to segregate the dataset into two sub-segments: a training subset and a testing subset, with the latter accounting for 20% of the complete training data.

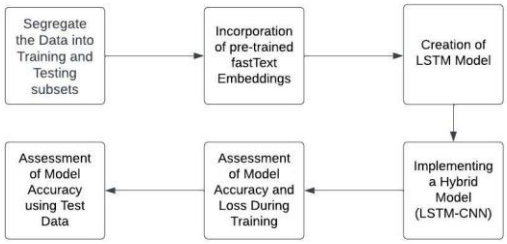


Fig. 4 Construction and Evaluation of the Models

2. *Step 2: Incorporation of pre-trained fastText Embeddings.*

In accordance with the previously stated objective, we aim to utilize the power of Transfer Learning by incorporating pre-trained word embedding from fastText. To achieve this, we import the fastText embedding into our workspace and construct an embedding matrix by mapping the vocabulary of our dataset to the pre-trained embedding. This is done by iterating over the words in our vocabulary and looking up their corresponding embedding in the pre-trained embedding. If some word in our vocabulary does not have a corresponding embedding in the pre-trained embedding, we can either randomly initialize its embedding or assign it a special out-of-vocabulary (OOV) token.

3. *Step 3: Creation of LSTM Model.*

With the train-set and train-set ready, it's time to choose a deep learning model for training. For a Natural Language Processing project like this one, the Long Short-Term Memory (LSTM) model is deemed ideal. LSTMs are similar to Recurrent Neural Networks (RNNs), However, they differ in one significant aspect - their hidden layer updates are substituted with memory cells, making them superior in detecting and revealing long-term relationships in data, which is critical for sentence structures.

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 200)]             0

 embeddings (Embedding)      (None, 200, 300)          44675400

 lstm_layer (LSTM)           (None, 200, 40)           54560

 global_max_pooling1d (Globa  (None, 40)               0
 lMaxPooling1D)

 dropout (Dropout)           (None, 40)                0

 dense (Dense)               (None, 30)                1230

 dropout_1 (Dropout)         (None, 30)                0

 dense_1 (Dense)             (None, 6)                 186

=================================================================
Total params: 44,731,376
Trainable params: 55,976
Non-trainable params: 44,675,400
_________________________________________________________________
```

Fig. 5 Model summary of LTSM Model

- To aid in hyperparameter tuning and model evaluation, the Talos library was imported. Using the "Scan" function, a GridSearchCV was carried out, resulting in the determination of the best parameters for achieving the highest accuracy.
- Afterwards, the LSTM model was configured with the best hyperparameters and required number of layers. Then the model was then compiled and trained utilizing both the testing and training set.

4. *Step 4: Implementing a Hybrid Model (LSTM-CNN).*

LSTMs in deep learning models are proficient in maintaining the features of historical information in extended sequences of text, while CNNs excel at extracting local features from the text. By integrating these two conventional neural network architectures, we can take advantage of their collective abilities. Consequently, we adopted the LSTM-CNN hybrid model to measure its performance compared to any other deep-learning models and identify the most effective one for our project.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:25:02 UTC from IEEE Xplore. Restrictions apply.

```
Model: "model_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_2 (InputLayer)        [(None, 200)]             0

 embeddings (Embedding)      (None, 200, 300)          44675400

 lstm_layer (LSTM)           (None, 200, 50)           70200

 conv1d (Conv1D)             (None, 200, 64)           9664

 max_pooling1d (MaxPooling1D  (None, 66, 64)           0
 )

 global_max_pooling1d_1 (Glo  (None, 64)               0
 balMaxPooling1D)

 batch_normalization (BatchN  (None, 64)               256
 ormalization)

 dense_2 (Dense)             (None, 40)                2600

 dropout_2 (Dropout)         (None, 40)                0

 dense_3 (Dense)             (None, 30)                1230

 dropout_3 (Dropout)         (None, 30)                0

 dense_4 (Dense)             (None, 6)                 186

=================================================================
Total params: 44,759,536
Trainable params: 84,008
Non-trainable params: 44,675,528
```

Fig. 6 Model summary of LTSM-CNN Model

Just like in the previous step, we utilized the "Talos" library to determine the optimal hyperparameters for the hybrid model. This Python library enables efficient exploration of hyper parameter space via randomized hyper parameter optimization. After completion of the hyper parameter optimization process, we analyzed the outcomes and picked the hyper parameters that resulted in the optimal performance metric on the testing set. This required evaluating the results of each hyper parameter combination and selecting the one that provided the highest accuracy. Subsequently, we trained the hybrid model on the complete training set utilizing this set of hyper parameters.

Finally, we trained the hybrid model using the testing and training sets. This involved providing the model with the preprocessed input data, the embedding matrix, and the hyper parameters obtained from Talos. The model was trained via backpropagation to minimize the loss function, this process entails computation of the gradients of the loss function with respect to the model parameters and using these gradients to adjust the parameters in the direction that decreases the loss.

When training the model using the training subset and testing it with the test set, it helps ensure that the model isn't overfitting to the training data and can generalize well to new, not previously seen data. Once the model has been trained, we move to the next step of assessment of calculating the accuracy of the model and loss during the training process.

5. *Step 5: Assessment of Model Accuracy and Loss During Training.*

Now that both of our deep-learning models have completed their training, it's important to utilize a visualization technique to demonstrate the accuracy and loss values of the models throughout their training phase. Normally, during the training process, the accuracy of a deep-learning model should rise as its loss value falls with the increase in the total count of epochs. The model quality and the success of its training can be effectively evaluated through this approach. We can observe the trends in accuracy and loss values for each epoch in the accompanying images.

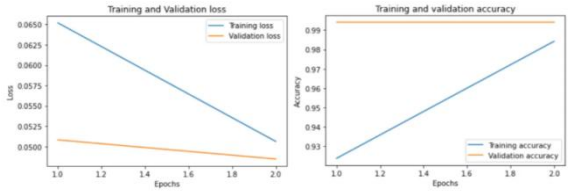


Fig. 7 For 2 epochs the LSTM model's Accuracy and Loss values.

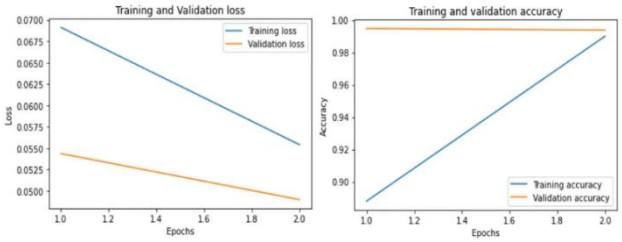


Fig. 8 For 2 epochs the LSTM-CNN model's Accuracy and Loss values.

6. *Step 6: Assessment of Model Accuracy using Test Data*

Having received encouraging results from the accuracy and loss evaluations, we can now proceed to gauge the quality of our deep learning models by testing them against the test data. This test data was procured from Kaggle and Similar data pre-processing and feature engineering steps were employed for both of the test and training data sets. With the now available processed test data, we can use the "predict" function to create the outputs for the inputs given from the test data.

In paper, Z. Zhai. (2022) [19] proposes a BERT-based deep learning method for rating the severity of toxic comments in online social media. The proposed method uses BERT-based backbones (RoBERTa and DeBERTa) and the Jigsaw dataset is used for training. The paper demonstrates the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:25:02 UTC from IEEE Xplore. Restrictions apply.

effectiveness of using deep learning and BERT-based models for identifying and rating toxic comments. Our method gives a better accuracy than the BERT based models proposed in the paper.Comparison of the accuracy of our proposed models with the ones presented in this paper can be shown as following:

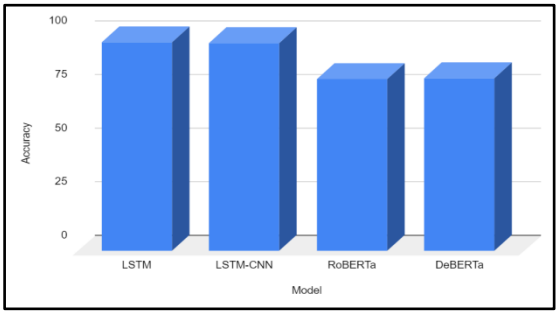


Fig. 9 Graphical representation of Model-Accuracy

TABLE 2 MODEL RESULTS

Model Accuracy				
Year	*Author*	*Model Name*	*Accuracy*	*Remarks*
2022	ZiyuZhai	RoBERTa	80.43	BERT is a pre-trained was designed primarily for transfer learning and fine-tuning specific tasks
		DeBERTa	80.56	
2023	Our proposed model	LTSM	97.40	LSTM performs better for sequential data and capturing long-term dependencies.

VI. CONCLUSION

In this paper, two approaches have been developed i.e. LSTM and hybrid LSTM-CNN models. After analyzing the accuracy achieved by the two different models, it can be declared that the conventional LSTM model surpasses the hybrid model. The LSTM model achieves the accuracy score of 97.40% while the LSTM-CNN model achieves the accuracy score of 97.07%. The traditional LSTM model appears to be the most appropriate choice for the Toxic Comment Classification task, given that the hybrid model's performance is only slightly lower than that of the traditional deep-learning model.

Future work could involve exploring multimodality and other types of hybrid deep learning models. Additionally, improving the data preprocessing techniques and exploring the use of additional features or different data augmentation techniques may further enhance the model's performance.

REFERENCES

[1] Bi-Directional LSTM Model for Question Answering on Knowledge Bases" by Zhengbao Zhang, Qi Liu, Huanbo Luan, and Maosong Sun, published in the Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018).

[2] Chen, T., Zhang, Y., & Chen, Y. (2019). A self-attentive model for toxic comment classification. In proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 1786-1795).

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

[4] Sun, B., Li, Y., Kang, Q., & Yang, W. (2020). Fine-tuning BERT for text classification of toxic comments. Journal of Ambient Intelligence and Humanized Computing, 11(7), 2703-2711.

[5] Qian, C., Liu, Y., & Yao, L. (2020). Hierarchical model for text classification with character-level and word-level embeddings. IEEE Access, 8, 134376-134387.

[6] Zhang, Y., Shen, Y., & Li, Y. (2021). Multi-task learning for toxic comment classification and user identification. Frontiers in Psychology, 12, 609925.

[7] Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. In Proceedings of the 11th International AAAI Conference on Web and Social Media (pp. 512-515).

[8] Waseem, Z., &Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. In Proceedings of the NAACL Student Research Workshop (pp. 88-93).

[9] Patil, S., Gugale, S., & Rode, M. (2018). Comparative analysis of LSTM and CNN-LSTM for sentiment analysis. In Proceedings of the 2nd International Conference on Communication and Electronics Systems (ICCES) (pp. 1105-1110).

[10] Xu, S., & Xu, R. (2019). Comparison of CNN and LSTM for Chinese text classification. In Proceedings of the 3rd International Conference on Education and Multimedia Technology (ICEMT) (pp. 117-121).

[11] T. Yao, Z. Zhai and B. Gao, "Text Classification Model Based on fastText," 2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS), Dalian, China, 2020, pp. 154-157, doi: 10.1109/ICAIIS49377.2020.9194939.

[12] Javed, F., Awan, S., Khan, M. A., & Tariq, U. (2019). A review of deep learning based approaches for hate speech detection. Journal of Ambient Intelligence and Humanized Computing,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:25:02 UTC from IEEE Xplore. Restrictions apply.

10(12), 4659-4673.
[13] Nobata, C., Tetreault, J., Thomas, A., Han, B., & Diab, M. T. (2016). Abusive language detection in online user content. In Proceedings of the 1st Workshop on Abusive Language Online (pp. 19-23).
[14] https://en.wikipedia.org/wiki/Long_short-term_memory
[15] Manchanda, C., Rathi, R., & Sharma, N. (2019). Traffic Density Investigation & Road Accident Analysis in India using Deep Learning. 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). doi: 10.1109/icccis48478.2019.8974528
[16] J. Zhang, Y. Li, J. Tian and T. Li, "LSTM-CNN Hybrid Model for Text Classification," 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 2018, pp. 1675-1680, doi: 10.1109/IAEAC.2018.8577620.
[17] Z. Zhai, "Rating the Severity of Toxic Comments Using BERT-Based Deep Learning Method," 2022 IEEE 5th International Conference on Electronics Technology (ICET), Chengdu, China, 2022, pp. 1283-1288, doi: 10.1109/ICET55676.2022.9825384
[18] Zhu, J., Xie, Y., Shen, J., & Hu, X. (2021). Dual-channel attention-based convolutional neural network for hate speech detection. Knowledge-Based Systems, 220, 106976.
[19] Zhai, Z. (2022). A BERT-based deep learning method for rating the severity of toxic comments in online social media. Applied Soft Computing, 117, 107978.
[20] Rustagi, A., Manchanda, C., Sharma, N., & Kaushik, I. (2020). Depression Anatomy Using Combinational Deep Neural Network. Advances in Intelligent Systems and Computing International Conference on Innovative Computing and Communications, 19-33. doi:10.1007/978-981-15-5148-2_3.
[21] Grover, M., Sharma, N., Bhushan, B., Kaushik, I., & Khamparia, A. (2020). Malware Threat Analysis of IoT Devices Using Deep Learning Neural Network Methodologies. Security and Trust Issues in Internet of Things, 123-143. doi:10.1201/9781003121664-6
[22] Sharma, A., Sharma, N., Kaushik, I., Kumar, S., &Khatoon, N. (2020). Predictive analysis of type 2 diabetes using hybrid ml model and iot. IoT Security Paradigms and Applications, 303-320. doi:10.1201/9781003054115-14.
[23] N. Sharma, I. Kaushik, B. Bhushan, C. K. Dixit et al., “Cryptocurrency Revolution: Bitcoin Time Forecasting &Blockchain Anomaly Detection,” in Blockchain Technology in Healthcare Applications. CRC Press, 2022, pp. 61–85.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:25:02 UTC from IEEE Xplore. Restrictions apply.

Water Air Soil Pollut (2023) 234:520
https://doi.org/10.1007/s11270-023-06548-6

Exploring the Impacts of HDPE Microplastics on Growth and Physiological Behavior of *Brassica juncea* (Mustard Plant)

Megha Bansal · Deenan Santhiya · Jai Gopal Sharma

Received: 24 May 2023 / Accepted: 22 July 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract In today's world, plastic pollution has become a big environmental concern. As a result, smaller plastic pieces, also known as microplastics (MPs), have received a lot of attention in recent years. The majority of study has focused on microplastic pollution in aquatic ecosystems and related shorelines. However, influence of microplastics (MPs) in the soil environment and on agroecosystem is minimal. To investigate the problem, this study is aimed to observe the effects of MPs in soil using *Brassica juncea* as a model plant. Because of its high biomass capacity, *Brassica juncea* can store considerable level of contaminants in its tissues. As a result, *B. juncea* was subjected to two microplastics, HDPE_MPs and HDPE_beads, followed by analysis of biometrical parameters, physio-biochemical features, and morphological analyses. Plants subjected to 20 g beads had the lowest concentrations of phenolic content, chlorophyll content, and proline content compared to other treated samples. Furthermore, fluorescence and confocal microscopy demonstrated the transfer of microplastics throughout plant roots and leaves, showing their potential for plant injury. To the best of our knowledge, this study focuses on the transfer of MPs from root to leaves of *Brassica juncea* for the first time. A method for microplastic uptake from plant roots to leaves has also been postulated. It merits further examination in future studies and provides new insight into the phytoaccumulation of MPs in soil.

Keywords Agroecosystem · Biometrical parameters · Microscopy · Physio-biochemical features · Phytoaccumulation

1 Introduction

The accumulation of emerging contaminants (ECs) in agricultural ecosystems is one of the main concern in environment in today's scenario (Taheran et al., 2018). The behavior, fate and ecological impacts of ECs has led to inadequate management and loss of biodiversity (Lodeiro et al., 2019). Plastic materials with small particle sizes, such as microplastics (MPs) (plastics with size < 1000 μm) and nanoplastics (NPs) (with particle size less than 1 μm), are of particular interest because of ecotoxicity and health hazards (Bermúdez & Swarzenski, 2021). Despite occurrence of microplastics by degradation, they are

M. Bansal · J. G. Sharma
Department of Biotechnology, Delhi Technological University, Delhi, India
e-mail: megha_2k21phdbt01@dtu.ac.in

J. G. Sharma
e-mail: sharmajaigopal@dce.ac.in

D. Santhiya (✉)
Department of Applied Chemistry, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi 110042, India
e-mail: deenan.santhiya@dce.ac.in

Published online: 29 July 2023

also incorporated as specific constituents in many products used in daily life. These microplastics bioaccumulate within the soil and have adverse effect on plant growth and development (Xu et al., 2020). For this reason, concern on impact of microplastics on plant performance, along with soil microbes and activity in soil has been the study for research (de Souza Machado et al., 2019; Ng et al., 2018).

Many recent studies have focused on understanding the impacts of various contaminants in soil and plants. For example, the leaves of *Prunus laurocerasus* L. obtained from diverse climatic habitats showed change in morphological characteristics due to different environmental conditions (YÏğit et al., 2018). Another study observed distinguished characteristics in various plant species exposed to magnesium, a major macronutrient for plants. Traffic density determined variations in magnesium concentrations in leaves of *Ficus bengalensis*, *Ziziphus mauritiana*, *Conocarpus erectus*, and *Azadirechta indica* species (Çetïn & Jawed, 2021). Accumulation of traffic-related heavy metals like aluminium, cadmium, iron, manganese, nickel, lead, and zinc in plants could be determined by using *Rosmarinus officinalis* L. (rosemary) as a model plant. This plant was mostly used in slopes and refuge along highway junctions and could act as a great tool for establishing traffic-related pollution in urban parts (Bozdogan Sert et al., 2019). Concentrations of heavy metals in soil can also increase on being accumulated with other contaminants like microplastics. For example, the uptake and inhibitory effects of cadmium on maize plants could be observed on being treated with HDPE microplastics. Different doses of microplastics induced phytotoxicity, change cadmium bioavailability and plant performance (Wang et al., 2020). Besides metal and other contaminant pollution in soil, microplastics could also have direct impacts on agroecosystem. For example, plastic mulch film significantly impact wheat growth, bacterial ecology and volatiles in the rhizosphere (Qi et al., 2020). Polylactic acid (PLA) microplastics severely decreased seedling height and prevented seed germination (Boots et al., 2019). In addition to plastic mulch film, scientists discovered that polystyrene-beads had effects on growth of plants on the basis of size, concentration, and that diameter may be a primary factor in determining whether micro/nanoplastics may permeate the tissues of plants. Regarding 100 nm polystyrene-beads, in addition to increasing wheat biomass (Lian et al., 2020), they might concentrate in the roots of *Vicia faba* and obstruct cell wall pores, which would impede delivery of plant nutrients (Jiang et al., 2019). According to a study, 10 mg kg^{-1} polystyrene-beads might shorten roots, whereas 100 nm enhanced seedling height (Ren et al., 2021). These findings demonstrated the buildup and movement of microplastics in the soil–plant system, highlighting the potential for microplastics to reach the food chain (Su et al., 2019). According to Dissanayake et al. (2022), applying agricultural sewage sludge alone leads to a large proportion of plastic residues into agricultural soils. Its input is predicted to be between 63,000 and 430,000 tonnes of microplastics per year in EU and North American farmlands, and between 2800 and 19,000 tonnes per year in Australian agroecosystems (Dissanayake et al., 2022). As per the study by Tun et al. (2022), 1411 pieces/kg of MPs were found in the Indian dumping soils which is just second to the highest pollution reported in Cambodia (Tun et al., 2022). Most of the microplastics that act as dominant polymers in soil include polyethylene, polyethylene terephthalate and polypropylene.

The major aim of present study is understanding uptake, accumulation, and translocation of high-density polyethylene (HDPE) microplastics along with their adverse effects on terrestrial plant. HDPE constitutes a large proportion of environmental pollution among all the microplastics (Awasthi et al., 2017). To facilitate this mechanism, *Brassica juncea*, commonly termed as Indian Mustard, was used as a model plant to assess sites of absorption, uptake, and accumulation within the plant. *Brassica juncea* is particularly useful for phytoremediation as it can accumulate high levels of heavy metals (lead, nickel, cadmium, mercury and selenium) in their tissues, a process called as phytoextraction (Rathore et al., 2019). *Brassica juncea* produce compounds called glucosinolates, which are broken down by enzymes to release toxic isothiocyanates. These isothiocyanates form complexes with heavy metals in the soil, where they are then absorbed by plant roots and stored in their tissues. This means they can quickly establish themselves in contaminated soil and start removing pollutants (Diwan et al., 2008). *Brassica juncea* has a high biomass, which means it can accumulate huge quantities of pollutants in tissues (Goswami & Das, 2015). Also, *Brassica juncea* has a deep root system that allows them to access pollutants that may

be located deep in soil. Overall, the combination of these characteristics makes *Brassica juncea* an effective plant for phytoremediation of soil. Therefore, it is considered as the most ideal plant to study potential for microplastics remediation in soil. To our knowledge, this study is the first to observe impacts of microplastics not only on roots, but also on leaves and shoots of plants. Biochemical analysis on roots, shoots and leaves has been demonstrated to provide further evidence of microplastics intake in plants. Finally, mechanism of uptake of microplastics by plants highlighting different pathways is briefed to observe phytoaccumulation and identify the possibilities for soil remediation of microplastics. Therefore, current study builds on examining the impacts of two different kinds of high-density polyethylene (HDPE) microplastics, namely HDPE_MPs and HDPE_beads, on *Brassica juncea*. To study the impact and accumulation of HDPE_MPs and HDPE_beads on the plant, biometrical parameters such as height and biomass was observed. Additionally, the amount of chlorophyll in leaves was used to determine the photosynthetic pigment. By analyzing the phenolic and proline concentrations in various plant sections, biochemical composition was estimated. An examination of the plant's morphology using fluorescence and confocal microscopy was done to see where microplastics were being absorbed. Finally, the mechanism of accumulation of MPs within plant roots and its translocation to aerial parts was highlighted to further our understanding of MPs impact in plants and soil.

2 Material And Methods

2.1 Experimental Design

2.1.1 Soil

To study impacts of different size microplastics on *Brassica juncea* (mustard seeds) in an open environment, a triplicate study was conducted. Pots containing equal concentrations of soil and mustard seeds with varying amounts of microplastics was used for the study. We harvested the mustard seeds after a week to determine the effects of our experiments on vegetative and reproductive growth. The sandy soil used in this study was obtained from an agricultural land in Ghaziabad, Uttar Pradesh, India, at 28° 39′ 14.1588" N and 77° 26′ 42.8784" E. The soil was composed of sand, silt, and clay with moderate amounts of organic matter. The air-dried soil was sieved with a 2 mm steel sieve before use.

2.1.2 Synthesis of Microplastics

In this experiment, two forms of microplastics were used: (1) high-density polyethylene (HDPE) microplastics and (2) high-density polyethylene (HDPE) beads. A method described by Crespy and Landfester, 2007, was used to prepare the HDPE microplastics (HDPE_MPs) and beads (HDPE_beads). A solution composed of 1 g of HDPE powder and 20 ml of xylene was mixed using a magnetic stirrer for 1 h until completely dissolved. The HDPE solution was then gently added to 100 ml of deionized water while maintaining the sonication at 70% amplitude (Branson W450 Digital sonicator, tip size 6.5 mm) for 30 s under ice cooling. The resulting solution was centrifuged, rinsed with water and ethanol, then air-dried before storage.

2.1.3 Mustard Seeds and Pots

Brassica juncea (mustard seeds) were obtained from Indian Institute of Agricultural Research (IARI), Pusa, New Delhi. The seeds were surface sterilized with 0.02% sodium hypochlorite (NaOCl) before immersion in 70% ethanol (Fig. 1: (A) *Brassica juncea* (mustard seeds) and (B) Germination done in complete dark). After sterilization, the seeds were rinsed several times with distilled water. The seeds were grown on tissue overnight before being planted in organic soil in November 2021. Pots were irrigated twice a week at first, then once every two days during seedling emergence in January and February of 2022. Because of the increase in temperature in February, the frequency of irrigation was increased. NPK was applied to the soil in each pot based on the mustard NPK requirement of 100, 20, and 60 kg/ha.

The pot used in the experiment was 20 cm long, 10 cm in diameter at the bottom, 13 cm in diameter at the top and had a volume of 2 l. We used a factorial experimental design. Furthermore, three control treatments with no microplastic residues were investigated. The experiment included 12 treatments performed in triplicate, as well as four independent pots containing tagged microplastics for imaging.

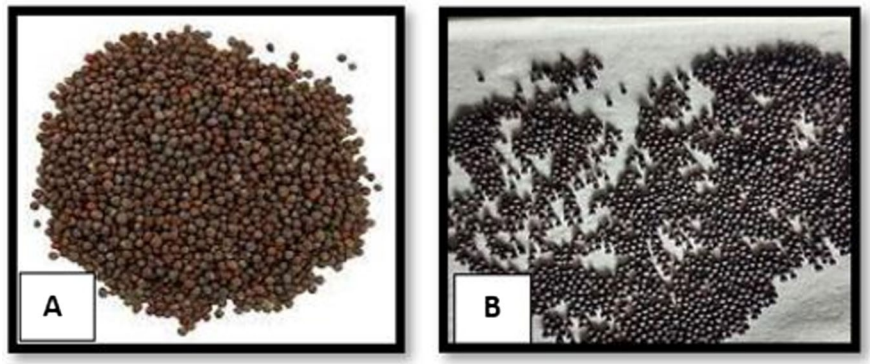


Fig. 1 (**A**) Brassica juncea (Mustard seeds) and (**B**) Germination done in complete dark

Each treatment was replicated three times, and total 36 pots of *B. juncea* seeds were grown. The mean of three potted plants was used to describe the study's findings.

2.2 Experimental Set-up and Climatic Conditions for Growth of Plant

2.2.1 Setting Up

Each pot constituted 2.5 kg of sieved soil and various concentrations of microplastics (except the three control treatments without microplastics) along with 150 ml of water. Prior to filling each pot with this mixture, a piece of geotextile was placed at the bottom of each pot to allow free circulation of air and water. After all the pots were filled, the soil moisture was uniformly set to 15%, which corresponds to the water capacity of the soil in the field. Before sowing *Brassica juncea* (mustard seeds), let sit for a week in each pot. Each container contained 12 g of litter (12.08 ± 0.06 g) and was sprayed with water to keep the litter moist.

2.2.2 Mustard Cultivation

Each pot contained 10–12 seeds, and post 2 weeks, 6–7 seedlings were selected from each pot for testing. The temperature was set at 15–16 °C during the day and 12 °C at night, with a photoperiod (14/10 h), a light intensity of 300 µmol m^{-2} s^{-1} and a relative humidity of 70% for day and night. The pots were watered weekly with tap water and soil moisture was adjusted to between 12 and 18% by weight. Pots were randomly placed in the climatic chamber and rotated each after two weeks.

2.2.3 Microplastic Tagging with Nile Red

For imaging, HDPE microplastic particles were labeled with Nile red as previously described (Karakolis et al., 2019). The dried microplastic particles were placed in an aqueous deionized (DI) solution of 100 µg/ml Nile red at a concentration of 50 mg of plastic particles per 10 ml of solution. This solution was prepared by dissolving 1 mg of Nile Red in 1 ml of acetone and the resulting solution was added to 10 ml of deionized water. The vials were left in the dark for two hours, rinsed three times with water, centrifuged and stored in deionized water for later use. Based on known staining procedures for microplastics, the Nile Red staining procedure was selected, including concentration of the Nile Red solution (Maes et al., 2017).

2.3 Measurements of Growth Parameters

2.3.1 Biometrical Analysis

Plant heights were measured using a steel tape measure on a regular basis from the 14th day after seeds were sown until the 90th day. Three months after planting, plants were divided into shoots and roots. Biometrical analysis provides the overall growth, biomass, and length of roots, shoots and leaves of plants after being exposed to different concentrations of a contaminant (Pricinotto et al., 2019). In this study, different concentrations of MPs were used as contaminant to observe the growth pattern in plants. At a span of three months from sowing of seeds, root and shoot height was measured and a mean of all three triplicates were used to determine the results of study. Similarly, root and shoot biomass was weighed

for each treated sample to determine mean difference with respect to control.

2.3.2 Chlorophyll Content

To determine photosynthetic pigment in leaves of plant, chlorophyll estimation was done. Relative chlorophyll content in plant leaves was measured and recorded as per the process defined by (Hong et al., 2012) using UV–Vis spectrophotometer (Biospectrophotometer, USA). Fresh leaves were obtained from each pot to determine the chlorophyll level. For each treated and control sample, 0.5 g of leaves were weighed from each pot. 10 ml of 80% acetone was added to chopped and homogenized leaves to make them transparent. The extract was centrifuged at 2500 rpm for five minutes. The resulting supernatant was diluted with 9 ml of 80% acetone before being measured with a UV–Vis Spectrophotometer at 663 nm and 644 nm. Total chlorophyll content in microplastic treated and control samples was evaluated using Mackinney's work and Arnon equations –

$$Chl_a = 12.7A_{663} - 2.69A_{645}; Chl_b = 22.9A_{645} - 4.68A_{663}$$
$$\text{Total chlorophyll} = Chl_a + Chl_b$$

2.4 Biochemical Analysis

2.4.1 Phenolic Content

A modified Folin–Ciocalteau test with gallic acid as the standard was used to assess total phenolic content (Ainsworth & Gillespie, 2007). 1 ml of plant extract was combined with 5 ml of Folin Ciocalteau's reagent after 1.5 ml of 20% Na_2CO_3 was added (diluted 1:10 with distilled water). Color development was accomplished by incubating the test tubes in dark for 30 min at room temperature, followed by measuring the absorbance at 765 nm. The total phenolic content of the sample was estimated as mg of dry mass equivalents of gallic acid. (GAE) mg^{-1}.

2.4.2 Proline Content

Modified ninhydrin chromogenic techniques were used to measure the proline content (Zhang et al., 2013). A glass tube containing freshly harvested roots (0.2 g) was filled with 5 ml of 3% sulfosalicylic acid. The glass tube was incubated for 10 min in a 100 °C water bath. 2 ml of the filtrate was digested in another glass tube after 4 ml of chromogenic solution (2 ml of 2.5% ninhydrin and 2 ml of glacial acetic acid) was added to the filter. The glass tube was then immersed for 30 min in a 100 °C water bath. Further, to stop the reaction, glass tube was submerged in an ice bath. 5 ml of toluene were placed in the glass tube, vortexed, and then allowed to stand. Using a spectrophotometer, the toluene layer's absorbance was observed at 520 nm.

2.5 Morphological Analysis

2.5.1 Fluorescence Microscopy

Fresh roots and leaves were removed from *Brassica juncea* plant and cleansed with deionized water. The roots were sectioned and placed on a glass slide with a few drops of clean water. The glass slide was then gently squeezed to flatten the pure water-covered root, ensuring that no air bubbles formed between the glass slide and cover slip. A fluorescent microscope was used to view each sample.

2.5.2 Confocal Microscopy

Fresh roots and leaves with tagged microplastics were picked out and cleansed with deionized water. On a glass slide with a few droplets of distilled water, cross-sections of roots were exhibited. Furthermore, to ensure no air bubbles between glass slide and cover slip, it was gently pressed to flatten the clean water-covered root. With the use of a confocal microscope (Nikon Laser Scanning Confocal Microscope), each sample was examined to observe the tagged microplastics. Similar observations were performed for leaf cross-sections of plant.

2.6 Statistical Analysis

All experiments were done in triplicate. Statistical analysis of experimental data was performed using Origin2023 software, and analysis of variance (ANOVA) was performed in SPSS 21.0 with a p-value of 0.05. For triplicate samples ($n=3$), all values were expressed as mean 5% standard error.

3 Results and Discussion

3.1 Biometrical Parameters

3.1.1 Growth Response

On exposure of *Brassica juncea* to HDPE_MPs and HDPE_beads, significant difference in shoot biomass and root biomass could be observed in contrast to control plants (Fig. 2 (a)). Shoot biomass for control was 3.155 ± 0.15 g compared to the treated plants showing a decreasing trend. The least shoot biomass was observed for 20 g MPs in contrast to control sample indicating an adverse effect when exposed to microplastics. Similarly, root biomass also showed a declining trend on being exposed to HDPE_MPs and HDPE_beads after a span of three months (Fig. 2 (b)). Root biomass for control was 0.85 ± 0.04 g compared to the treated plants having significant difference in values.

The shoot and root height also showed a declining trend on being treated with HDPE_MPs and HDPE_beads (Fig. 2 (c) & (d)). The shoot height for control plant was 33.23 ± 1.66 cm compared to microplastics treated samples showing declining trend. Also, the root height for control sample was 5.6 ± 0.28 cm compared to microplastics exposed samples showing a decreasing trend. Figure 2: Influence of HDPE_MPs and HDPE_beads on plant samples by observing growth parameters: (a) Shoot Biomass; (b) Root Biomass; (c) Shoot Height (The values are mean of three triplicates; Error bars denote 5% standard error).

3.1.2 Chlorophyll Content

The results showed that chlorophyll *a* (Chl *a*) was more easily impacted by varied microplastic concentrations than chlorophyll *b* (Chl *b*). Such a considerable variation in chl *a* and chl *b* content implies that total chlorophyll content in leaves of treated and control samples changed significantly (Fig. 3 (a) & (b)).

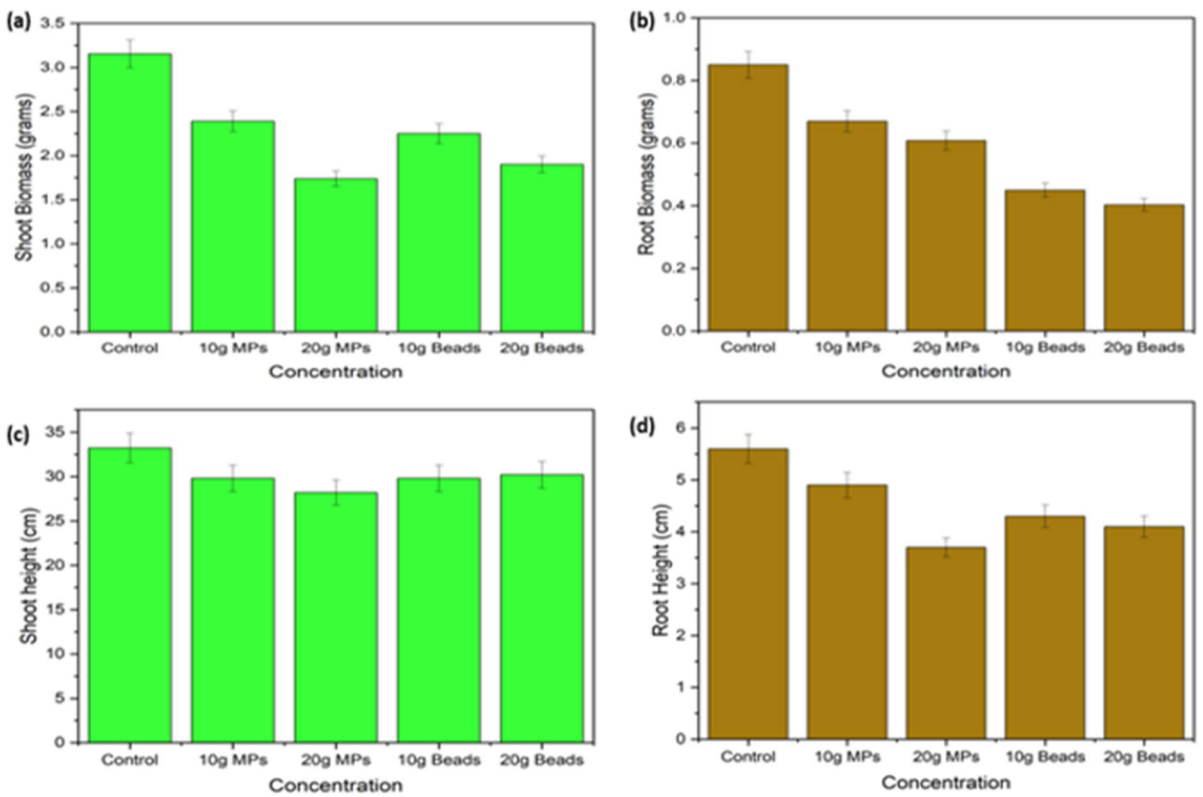


Fig. 2 Influence of HDPE_MPs and HDPE_beads on plant samples by observing growth parameters: (**a**) Shoot Biomass; (**b**) Root Biomass; (**c**) Shoot Height; (**d**) Root Height (The values are mean of three triplicates; Error bars denote 5% standard error)

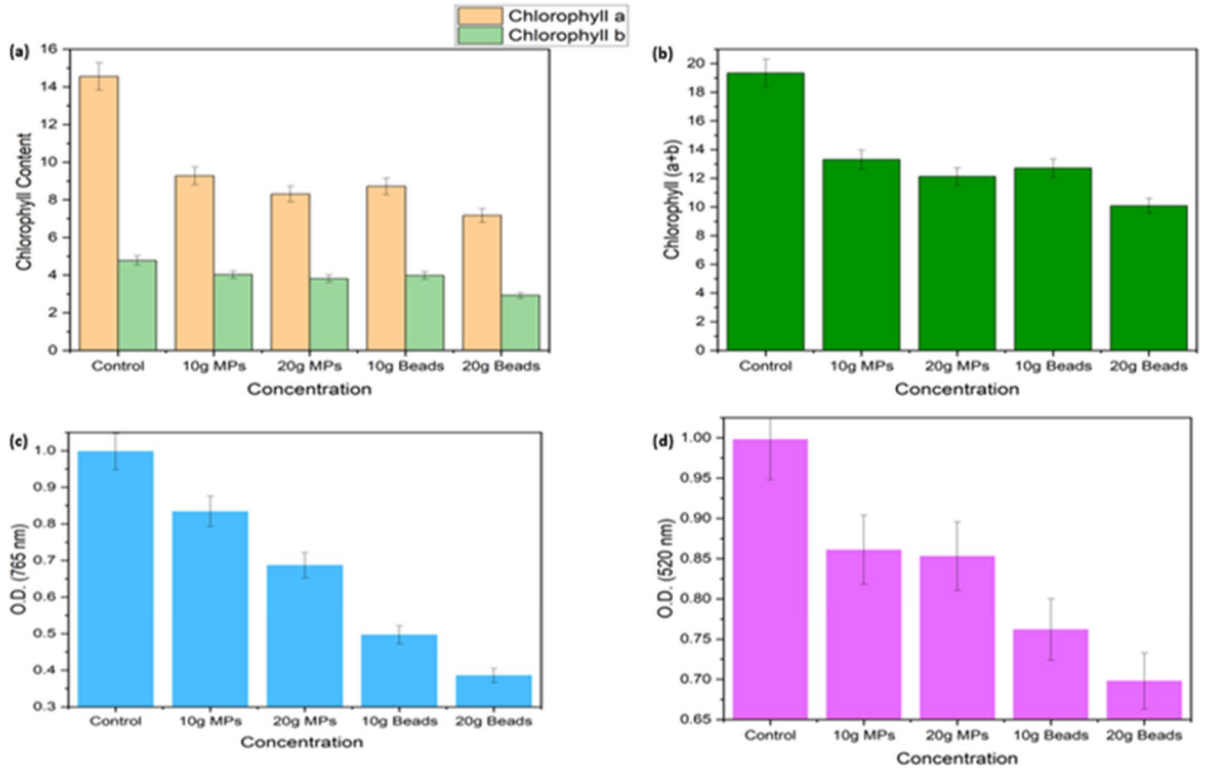


Fig. 3 Influence of HDPE_MPs and HDPE_beads on plant samples by observing growth parameters: (**a**) Chlorophyll a & b; (**b**) Total chlorophyll; (**c**) Phenolic content; (**d**) Proline content (The values are mean of three triplicates; Error bars denote 5% standard error)

Chlorophyll *a* is the primary pigment in the photosynthesis process, whereas chlorophyll *b* is an additional pigment that delivers energy to chlorophyll *a* (Khaleghi et al., 2013). An important metric for photosynthetic activity is the overall chlorophyll concentration (chl *a* + chl *b*), and variations in this value are a sign of stress in plants. Thus, it could be observed that chlorophyll *a* was comparatively less for microplastics treated plant samples indicating that total chlorophyll content in leaves showed greater inhibitory effect on exposure to microplastics.

3.2 Biochemical Analysis

3.2.1 Phenolic Content

The most common secondary metabolites found in plants are phenolics that serve as biological aid in plants, including defense against pathogens, protection from ultraviolet rays, pigmentation to draw pollinators, and defense against reactive oxygen species (Waśkiewicz et al., 2013). On being exposed to microplastics of different types, phenolic content in plants reduced showing a decreasing trend as observed in Fig. 3 (c). Compared to control having phenolic content of 0.998 ± 0.04 (GAE) mg^{-1}, phenolic content in plants treated with microplastics showed declining trend. The results indicate that incorporation of microplastics decreases phenolic content in plants thereby inducing stress. Figure 3: Influence of HDPE_MPs and HDPE_beads on plant samples by observing growth parameters: (a) Chlorophyll a & b; (b) Total chlorophyll; (c) Phenolic content; (d) Proline content (The values are mean of three triplicates; Error bars denote 5% standard error).

3.2.2 Proline Content

The current investigation demonstrated that *B. juncea* plants treated with HDPE_MPs and HDPE_beads had

a reduced proline content (Fig. 3 (d)). As proline is known to reduce oxidative stress, maintain osmotic balance, and regulate redox potential, increasing concentration of microplastics showed an opposite trend. On being exposed to MPs, content of proline gradually declined with 20 g beads showing lowest value of 0.698 ± 0.03 gg^{-1} compared to control with proline content of 0.998 ± 0.04 gg^{-1}. These values are indicative of the fact that proline is not able to reduce stress in *Brassica juncea* plants contaminated with microplastics.

The results presented in Fig. 2 (a) to (d) and Fig. 3 (a) to (d) are in good agreement and show the impact of microplastics on *Brassica juncea*. The decreasing trend observed in shoot biomass, root biomass, shoot height and root height ascertain the uptake of microplastics by plants and also provide evidence of plant contaminated with microplastics. Increasing concentration of microplastics in each of triplicate samples decreases the biometrical growth of *Brassica juncea*. Also, in contrast to chlorophyll *b*, chlorophyll *a* showed a major declining concentration on adding microplastics. Since, the major pigment involved in photosynthetic energy production is chlorophyll *a* (Khaleghi et al., 2013), its low content in microplastics treated samples indicates that total chlorophyll content gets significantly affected on addition of microplastics. Similarly, biochemical analysis conveys the impact of microplastics on *Brassica juncea* plants. As observed in Fig. 3 (c), phenolic content plays a major role as antioxidant agents that aid in free radical scavenging (Aryal et al., 2019). Decreasing concentration of phenolic agents on addition of microplastics justifies that plant have less redox potential thereby showing less antioxidant activity. Figure 3 (d) showing proline concentration signifies that microplastics addition induces stress in plants (Hayat et al., 2012).

3.3 Morphological Analysis

3.3.1 Fluorescence Microscopy

In this study, we show that *Brassica juncea* plant roots may absorb microplastics with sizes ranging between 5 to 10 µm from the surrounding soil (Fig. 4 (a) & (b)). We were able to find and see labelled microplastic particles embedded amid root cell structures using fluorescence microscopy. Fluorescing microplastic was found in exodermis, and vascular tissues, as well as in root hairs and outer epidermal layer, in addition to the inherent autofluorescence

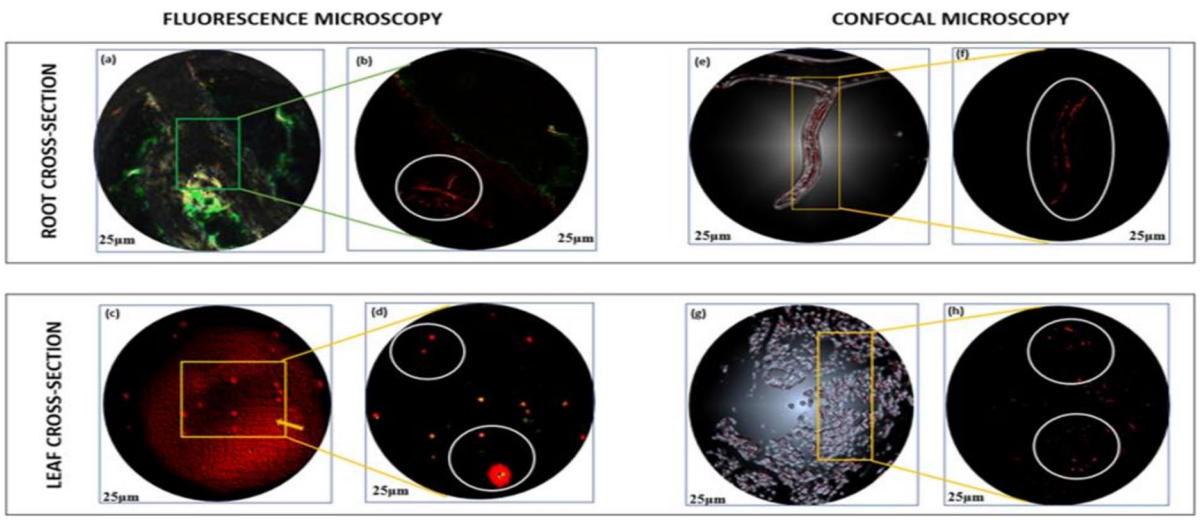


Fig. 4 Longitudinal cross-section of a Brassica juncea lateral root and leaf section in a three-month-old mustard plant after being exposed to a tagged microplastics. Clockwise from top left, flourescence microscopy shows root cross-sections (**a**) and (**b**), leaf sections (**c**) and (**d**), and confocal laser scanning microscopy shows root cross-sections (**e**) and (**f**), and leaf sections (**g**) and (**h**). Flourescing microplastic particles are indicated by white circles. All photographs were captured at magnifications of 40X and 100X and enhanced with contrast and brightness correction. Scales are displayed in the lower left and right corner of each image

of tree root. (Fig. 4(c) & (d)) provides evidence for microplastics incorporation in leaves of mustard plant as observed with red tagged MPs. Tagged microplastics could be observed in leaf sections and also induce structural changes in the morphology of leaf which is in line with our physiological and biochemical analysis (Sections 3.1. and 3.2.).

3.3.2 Confocal Microscopy

To decipher more significant findings of the study, confocal microscopy was performed to visualize tagged microplastics within the root lateral cross-sections and leaf parts. On visualization under red and green field at 40X and 100X magnification, it could be observed that microplastics were embedded in root hair segments and leaf internal veins at some points (Fig. 4 (e), (f), (g) & (h)). Because we could observe microplastics particles in inner root and leaf structures, our results suggest that micrometer sized microplastic can easily traverse from soil to root through crack-entry and apoplastic pathway, forming basis for entry to food chain. These results confirm microplastics uptake by *Brassica juncea* plants and also provide future insights for understanding the mechanism of action and impact on plants. Figure 4: Longitudinal cross-section of a *Brassica juncea* lateral root and leaf section in a three-month-old mustard plant after being exposed to tagged microplastics. Clockwise from top left, fluorescence microscopy shows root cross-sections (a) and (b), leaf sections (c) and (d), and confocal laser scanning microscopy shows root cross-sections (e) and (f), and leaf sections (g) and (h). Fluorescing microscopic particles are indicated by white circles. All photographs were captured at magnifications of 40X and 100X and enhanced with contrast and brightness correction. Scales are displayed in the lower left and right corner of each image.

To confirm the uptake of microplastics in root and leaf sections, morphological analysis using fluorescence and confocal microscopy is presented that is in line with the previous results. Biometrical parameters and biochemical analysis convey impact of microplastics on *Brassica juncea*. To provide further evidence for the same, fluorescence microscopy observed in Fig. 4 (a) to (d) signifies microplastics uptake in root and leaf cross-sections. Also, confocal microscopy shown in Fig. 4 (e) to (h), confirms the results and is in accordance with fluorescence microscopy observations.

3.4 Mechanism of Microplastics Uptake by Plants

Through this study, it could be observed that microplastics traverse from root hair sections to upper parts of plant like leaves and shoot. The results signify that microplastics have possible implications on *Brassica juncea* plant after being exposed. However, the mechanism of microplastic passage from soil to different plant parts is still in its infancy and needs more research. The apoplast and symplast routes are two pathways that microplastics can take to move through the plant from roots to the leaves (Su et al., 2019). Apoplast is the space outside plant cells, consisting of cell walls, intercellular spaces, and extracellular fluid. Microplastics can move through this space via diffusion, and uptake into the plant is thought to occur through the root epidermis and cortex as observed by morphological analysis in Fig. 4 (a), (b), (e) and (f). Once inside the apoplast, microplastics can move laterally along the root cell walls and through intercellular spaces to reach the xylem vessels (Roberts & Oparka, 2003). From xylem, microplastics are up taken by leaf parts and other tissues of plant as observed in Fig. 4 (c), (d), (g) and (h). The apoplast route is thought to be the primary pathway for larger microplastics to traverse the plant.

The symplast is the interconnected network of plant cells via plasmodesmata, which are small channels that allow for direct communication and transport of molecules between cells. Microplastics can enter the plant cells through endocytosis or other mechanisms and move through the cytoplasmic continuum from cell to cell via plasmodesmata (Raliya et al., 2016). This route is thought to be the primary pathway for smaller microplastics or those with a hydrophilic surface. The probable pathway for microplastic transport within the plant is described in Fig. 5: Mechanism of transport of microplastics from roots to different parts of plant. Some general mechanisms that are mostly proposed include:

a. Adhesion and penetration: Microplastics may adhere to the surface of roots and enter into root tissues by physical or chemical interactions (Nel et al., 2009). Above morphological analysis (Fig. 4) confirms the uptake of microplastics

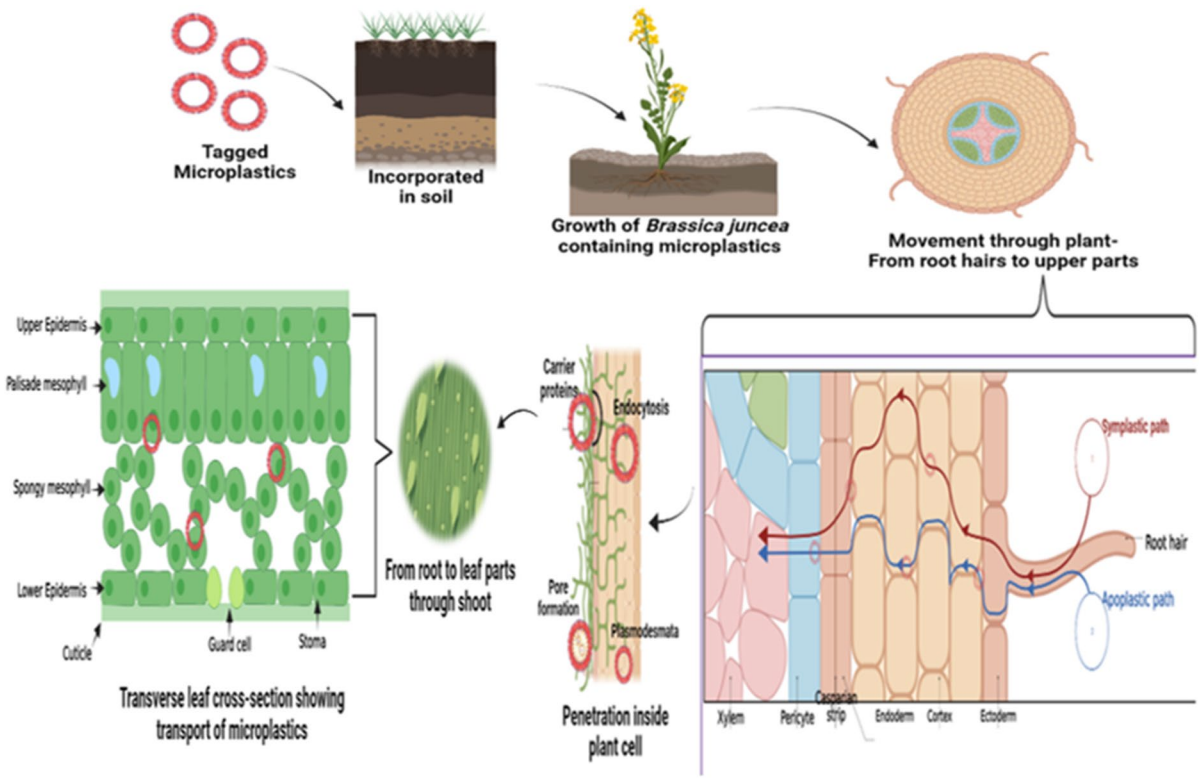


Fig. 5 Mechanism of transport of microplastics from roots to different parts of plants

from soil to root hairs and is in good agreement with the theoretical observations.

b. Endocytosis: Microplastics can traverse to plant cells through endocytosis, that encompasses formation of vesicles around the particles (Etxeberria et al., 2006). As observed in Fig. 4 (d), leaf sections show degradation of leaf surface due to accumulation of tagged microplastics. These results confirm the inflammation of leaf surface on being contaminated with microplastics.
c. Translocation: Once inside the root, microplastics may be transported across the root cortex and into the xylem vessels, which carry water and nutrients up to the leaves. This transport can cause microplastics uptake and occur through diffusion or active transport mechanisms (Schwab et al., 2016). The morphological analysis observed in Fig. 4 (a) to (h) shows that microplastics traverse from root hairs to leaf sections possibly through nutrients and water.
d. Accumulation in leaf parts: Once in xylem vessels, microplastics can be transported to the leaves, where they can accumulate in the leaf tissues. This accumulation can occur through transpiration, which is the loss of water through leaves, and subsequent concentration of microplastics in the leaf tissues. As observed in Fig. 4 (e) and (f), the leaf sections present accumulation of tagged microplastics at the surface and also degradation of a part due to accumulation.

The uptake of microplastics by plants can have various effects, including alterations in plant growth, development, and metabolism. This is well confirmed with the above findings stating impact of microplastics on biometry, biochemical analysis, and morphology. Furthermore, the presence of microplastics in edible plant tissues could pose potential risks to humans, and more research is required to understand their impacts on plants. Hence, the migration of MPs inside plants is really important as it gives an evidence on the accumulation and absorption of MPs on plant parts. Also, after accumulation and absorption of MPs, the phytoremediation potential can be

determined to provide sustainable solution for environmental cleanup.

4 Salient Findings of the Study

The results from overall study provide evidence for microplastics uptake by *Brassica juncea* and its impact on growth and development. The biometrical parameters and biochemical analysis signify that microplastics are transported from root to shoot parts of the plant as depicted by Figs. 2 and 3. Also, morphological analysis confirm uptake of microplastics from root to leaf sections of plant. The shoot biomass showed least concentration of 1.75 g on addition of 20 g MPs. Similarly, root biomass of 0.4 g was observed on addition of 20 g beads, that was least compared to all other concentrations. Also, the least shoot and root height was observed in plants treated with 20 g MPs showing shoot height at 27 cm and root height at 3.5 cm (Fig. 2). Similar to this, least chlorophyll concentration was observed in 20 g beads thereby showing least total chlorophyll content in microplastics exposed plants. Also, the antioxidant activity depicted by phenolic content and protection of plant against stress represented by proline content was least for 20 g beads exposed plants (Fig. 3). Morphological analysis confirmed the presence of microplastics uptake in root and leaf sections of plant. Confocal microscopy provides a rapid approach for visualization of MPs within plant parts (Li et al., 2020). The fluorescent dyes can generate stable emission signals that are easy to distinguish from the autofluorescence generated by plant tissues (Zhang et al., 2022). Thus, it is a rapid and efficient approach in detection of MPs within plant tissues (Ullah et al., 2021). The accumulation of tagged microplastics could also be observed in Fig. 4 (c) and (d) sections. These results portray uptake of microplastics not only by root but also by leaf sections of plant. Thus, *Brassica juncea* have been found to accumulate microplastics in their tissues, that could have significant impact on humans when consumed. This study also highlights uptake mechanism of microplastics from root to leaf sections as confirmed by morphological analysis. All the results present a significant finding providing evidence on impact on *Brassica juncea* plant after being exposed to different concentrations of microplastics. Further studies exploring phytoaccumulation of microplastics and remediation of soil need to be investigated for providing sustainability to agricultural ecosystem.

5 Conclusion

The inference from this study demonstrates absorption, uptake and phytotoxicity of microplastics to *Brassica juncea*, a vascular terrestrial plant for first time. Physiological and biochemical analysis reveals transport of MPs from root to leaf sections. Morphological analysis confirms the existence of MPs in various parts of plant as observed in vascular tissues of leaf sections and root hairs. Based on these findings, mechanism of transport of MPs from root to leaf sections through apoplast and symplast pathway is proposed. The accumulation of microplastics is evidenced in *Brassica juncea* plants providing a platform for future studies on remediation of soil in combination with other approaches to achieve optimal results. Nonetheless, it holds promise as a sustainable and environment friendly approach to addressing the issue of microplastic pollution.

Acknowledgements The authors like to thank Delhi Technological University for providing infrastructure and facility.

Author Contribution **Megha Bansal**: Investigation, Methodology, Writing—original draft. **Deenan Santhiya**: Conceptualization, Methodology, Writing—review & editing, Supervision, **Jai Gopal Sharma**: Funding acquisition, Resources, Supervision.

Data Availability All the data generated and analyzed during the current study is included in the manuscript.

Declarations

Financial Interests The authors have no relevant financial or non-financial interests to disclose.

Competing Interests The authors declare that they have no conflict of interest.

References

Ainsworth, E. A., & Gillespie, K. M. (2007). Estimation of total phenolic content and other oxidation substrates in plant tissues using Folin-Ciocalteu reagent. *Nature*

Protocols, 2(4), 875–877. https://doi.org/10.1038/nprot.2007.102

Aryal, S., Baniya, M. K., Danekhu, K., Kunwar, P., Gurung, R., & Koirala, N. (2019). Total phenolic content, flavonoid content and antioxidant potential of wild vegetables from Western Nepal. *Plants, 8*(4), 96. https://doi.org/10.3390/plants8040096

Awasthi, S., Srivastava, P., Singh, P., Tiwary, D., & Mishra, P. K. (2017). Biodegradation of thermally treated high-density polyethylene (HDPE) by Klebsiella pneumoniae CH001. *3 Biotech, 7*(5), 332. https://doi.org/10.1007/s13205-017-0959-3

Bermúdez, J. R., & Swarzenski, P. W. (2021). A microplastic size classification scheme aligned with universal plankton survey methods. *Methods X*, *8*, 101516. https://doi.org/10.1016/j.mex.2021.101516

Boots, B., Russell, C. W., & Green, D. S. (2019). Effects of microplastics in soil ecosystems: Above and below ground. *Environmental Science & Technology, 53*(19), 11496–11506. https://doi.org/10.1021/acs.est.9b03304

Bozdogan Sert, E., Turkmen, M., & Cetin, M. (2019). Heavy metal accumulation in rosemary leaves and stems exposed to traffic-related pollution near Adana-İskenderun Highway (Hatay, Turkey). *Environmental Monitoring and Assessment, 191*(9), 553. https://doi.org/10.1007/s10661-019-7714-7

Çetin, M., & Jawed, A. A. (2021). The chancing of Mg concentrations in some plants grown in Pakistan depends on plant species and the growing environment. *Kastamonu University Journal of Engineering and Sciences, 7*(2), 167–174.

Crespy, D., & Landfester, K. (2007). Preparation of nylon 6 nanoparticles and nanocapsules by two novel miniemulsion/solvent displacement hybrid techniques. *Macromolecular Chemistry and Physics, 208*(5), 457–466. https://doi.org/10.1002/macp.200600487

de Souza Machado, A. A., Lau, C. W., Kloas, W., Bergmann, J., Bachelier, J. B., Faltin, E., et al. (2019). Microplastics can change soil properties and affect plant performance. *Environmental Science & Technology, 53*(10), 6044–6052. https://doi.org/10.1021/acs.est.9b01339

Dissanayake, P. D., Kim, S., Sarkar, B., Oleszczuk, P., Sang, M. K., Haque, M. N., et al. (2022). Effects of microplastics on the terrestrial environment: A critical review. *Environmental Research, 209*, 112734. https://doi.org/10.1016/j.envres.2022.112734

Diwan, H., Ahmad, A., & Iqbal, M. (2008). Genotypic variation in the phytoremediation potential of Indian mustard for chromium. *Environmental Management, 41*(5), 734–741. https://doi.org/10.1007/s00267-007-9020-3

Etxeberria, E., Gonzalez, P., Baroja-Fernández, E., & Romero, J. P. (2006). Fluid phase endocytic uptake of artificial nano-spheres and fluorescent quantum dots by sycamore cultured cells. *Plant Signaling & Behavior, 1*(4), 196–200. https://doi.org/10.4161/psb.1.4.3142

Goswami, S., & Das, S. (2015). A study on cadmium phytoremediation potential of indian mustard, Brassica juncea. *International Journal of Phytoremediation, 17*(6), 583–588. https://doi.org/10.1080/15226514.2014.935289

Hayat, S., Hayat, Q., Alyemeni, M. N., Wani, A. S., Pichtel, J., & Ahmad, A. (2012). Role of proline under changing environments. *Plant Signaling & Behavior, 7*(11), 1456–1466. https://doi.org/10.4161/psb.21949

Hong, R., Feng, L., & Yan, X. (2012). Comparison on methods of chlorophyll extraction in flowering chinese cabbage. *Journal of Anhui Agricultural Sciences*. https://scholar.google.com/scholar_lookup?title=Comparison+on+methods+of+chlorophyll+extraction+in+flowering+chinese+cabbage&author=Ren+Hong%2C+Sanya+Science+and+Technology+Academy+for+Crop+Winter+Multiplication%2CSanya%28China%29&publication_year=2012. Accessed 5 February 2023

Jiang, X., Chen, H., Liao, Y., Ye, Z., Li, M., & Klobučar, G. (2019). Ecotoxicity and genotoxicity of polystyrene microplastics on higher plant Vicia faba. *Environmental Pollution, 250*, 831–838. https://doi.org/10.1016/j.envpol.2019.04.055

Karakolis, E. G., Nguyen, B., You, J. B., Rochman, C. M., & Sinton, D. (2019). Fluorescent dyes for visualizing microplastic particles and fibers in laboratory-based studies. *Environmental Science & Technology Letters, 6*(6), 334–340. https://doi.org/10.1021/acs.estlett.9b00241

Khaleghi, E., Arzani, K., Moallemi, N., & Barzegar, M. (2013). Evaluation of chlorophyll content and chlorophyll fluorescence parameters and relationships between chlorophyll a, b and chlorophyll content index under water stress in Olea europaea cv. Dezful. International Journal of Agricultural and Biosystems Engineering 6(8), 636–639. https://doi.org/10.1016/j.mex.2021.101516

Li, L., Luo, Y., Peijnenburg, W. J. G. M., Li, R., Yang, J., & Zhou, Q. (2020). Confocal measurement of microplastics uptake by plants. *MethodsX, 7*, 100750. https://doi.org/10.1016/j.mex.2019.11.023

Lian, J., Wu, J., Xiong, H., Zeb, A., Yang, T., Su, X., et al. (2020). Impact of polystyrene nanoplastics (PSNPs) on seed germination and seedling growth of wheat (Triticum aestivum L.). *Journal of Hazardous Materials, 385*, 121620. https://doi.org/10.1016/j.jhazmat.2019.121620

Lodeiro, C., Capelo, J. L., Oliveira, E., & Lodeiro, J. F. (2019). New toxic emerging contaminants: Beyond the toxicological effects. *Environmental Science and Pollution Research, 26*(1), 1–4. https://doi.org/10.1007/s11356-018-3003-1

Maes, T., Jessop, R., Wellner, N., Haupt, K., & Mayes, A. G. (2017). A rapid-screening approach to detect and quantify microplastics based on fluorescent tagging with Nile Red. *Scientific Reports, 7*(1), 44501. https://doi.org/10.1038/srep44501

Nel, A. E., Mädler, L., Velegol, D., Xia, T., Hoek, E. M. V., Somasundaran, P., et al. (2009). Understanding biophysicochemical interactions at the nano–bio interface. *Nature Materials, 8*(7), 543–557. https://doi.org/10.1038/nmat2442

Ng, E.-L., Huerta Lwanga, E., Eldridge, S. M., Johnston, P., Hu, H.-W., Geissen, V., & Chen, D. (2018). An overview of microplastic and nanoplastic pollution in agroecosystems. *Science of the Total Environment, 627*, 1377–1388. https://doi.org/10.1016/j.scitotenv.2018.01.341

Pricinotto, L. F., Zucareli, C., Ferreira, A. S., Spolaor, L. T., & Fonseca, I. C. D. B. (2019). Yield and biometric characteristics of maize submitted to plant population and

trinexapac-ethyl doses. *Revista Caatinga, 32*, 667–678. https://doi.org/10.1590/1983-21252019v32n311rc

Qi, Y., Ossowicki, A., Yang, X., Huerta Lwanga, E., Dini-Andreote, F., Geissen, V., & Garbeva, P. (2020). Effects of plastic mulch film residues on wheat rhizosphere and soil properties. *Journal of Hazardous Materials, 387*, 121711. https://doi.org/10.1016/j.jhazmat.2019.121711

Raliya, R., Franke, C., Chavalmane, S., Nair, R., Reed, N., & Biswas, P. (2016). Quantitative Understanding of Nanoparticle Uptake in Watermelon Plants. *Frontiers in Plant Science*, *7*. https://www.frontiersin.org/articles/10.3389/fpls.2016.01288. Accessed 18 October 2022

Rathore, S. S., Shekhawat, K., Dass, A., Kandpal, B. K., & Singh, V. K. (2019). Phytoremediation mechanism in Indian mustard (Brassica juncea) and its enhancement through agronomic interventions. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences, 89*(2), 419–427. https://doi.org/10.1007/s40011-017-0885-5

Ren, X., Tang, J., Wang, L., & Liu, Q. (2021). Microplastics in soil-plant system: Effects of nano/microplastics on plant photosynthesis, rhizosphere microbes and soil properties in soil with different residues. *Plant and Soil, 462*(1), 561–576. https://doi.org/10.1007/s11104-021-04869-1

Roberts, A. G., & Oparka, K. J. (2003). Plasmodesmata and the control of symplastic transport: Plasmodesmata and symplastic transport. *Plant, Cell & Environment, 26*(1), 103–124. https://doi.org/10.1046/j.1365-3040.2003.00950.x

Schwab, F., Zhai, G., Kern, M., Turner, A., Schnoor, J. L., & Wiesner, M. R. (2016). Barriers, pathways and processes for uptake, translocation and accumulation of nanomaterials in plants – Critical review. *Nanotoxicology, 10*(3), 257–278. https://doi.org/10.3109/17435390.2015.1048326

Su, Y., Ashworth, V., Kim, C., Adeleye, A. S., Rolshausen, P., Roper, C., et al. (2019). Delivery, uptake, fate, and transport of engineered nanoparticles in plants: A critical review and data analysis. *Environmental Science: Nano, 6*(8), 2311–2331. https://doi.org/10.1039/C9EN00461K

Taheran, M., Naghdi, M., Brar, S. K., Verma, M., & Surampalli, R. Y. (2018). Emerging contaminants: Here today, there tomorrow! *Environmental Nanotechnology, Monitoring & Management, 10*, 122–126. https://doi.org/10.1016/j.enmm.2018.05.010

Tun, T. Z., Kunisue, T., Tanabe, S., Prudente, M., Subramanian, A., Sudaryanto, A., et al. (2022). Microplastics in dumping site soils from six Asian countries as a source of plastic additives. *Science of The Total Environment, 806*, 150912. https://doi.org/10.1016/j.scitotenv.2021.150912

Ullah, R., Tsui, M.T.-K., Chen, H., Chow, A., Williams, C., & Ligaba-Osena, A. (2021). Microplastics interaction with terrestrial plants and their impacts on agriculture. *Journal of Environmental Quality, 50*(5), 1024–1041. https://doi.org/10.1002/jeq2.20264

Wang, F., Zhang, X., Zhang, S., Zhang, S., Adams, C. A., & Sun, Y. (2020). Effects of co-contamination of microplastics and Cd on plant growth and Cd accumulation. *Toxics, 8*(2), 36. https://doi.org/10.3390/toxics8020036

Waśkiewicz, A., Muzolf-Panek, M., & Goliński, P. (2013). Phenolic content changes in plants under salt stress. In P. Ahmad, M. M. Azooz, & M. N. V. Prasad (Eds.), *Ecophysiology and responses of plants under salt stress* (pp. 283–314). Springer. https://doi.org/10.1007/978-1-4614-4747-4_11

Xu, B., Liu, F., Cryder, Z., Huang, D., Lu, Z., He, Y., et al. (2020). Microplastics in the soil environment: Occurrence, risks, interactions and fate – A review. *Critical Reviews in Environmental Science and Technology, 50*(21), 2175–2222. https://doi.org/10.1080/10643389.2019.1694822

Yİğit, N., ÇetİN, M., & ŞevİK, H. (2018). The change in some leaf micromorphological characters of Prunus laurocerasus L. species by their habitat. *Turkish Journal of Agriculture - Food Science and Technology, 6*(11), 1517–1521.

Zhang, X., Huang, G., Bian, X., & Zhao, Q. (2013). Effects of root interaction and nitrogen fertilization on the chlorophyll content, root activity, photosynthetic characteristics of intercropped soybean and microbial quantity in the rhizosphere. *Plant, Soil and Environment, 59*(No. 2), 80–88. https://doi.org/10.17221/613/2012-PSE

Zhang, Z., Cui, Q., Chen, L., Zhu, X., Zhao, S., Duan, C., et al. (2022). A critical review of microplastics in the soil-plant system: Distribution, uptake, phytotoxicity and prevention. *Journal of Hazardous Materials, 424*, 127750. https://doi.org/10.1016/j.jhazmat.2021.127750

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Exploring Tidal Force Effects and Shadow Constraints for Schwarzschild-like Black Hole in Starobinsky-Bel-Robinson Gravity

Dhruv Arora,[1,*] Niyaz Uddin Molla,[2,†] Himanshu Chaudhary,[3,1,4,‡]
Ujjal Debnath,[2,§] Farruh Atamurotov,[5,6,7,¶] and G.Mustafa[8,9,**]

[1] *Pacif Institute of Cosmology and Selfology (PICS), Sagara, Sambalpur 768224, Odisha, India*
[2] *Department of Mathematics, Indian Institute of Engineering Science and Technology, Shibpur, Howrah-711 103, India,*
[3] *Department of Applied Mathematics, Delhi Technological University, Delhi-110042, India,*
[4] *Department of Mathematics, Shyamlal College, University of Delhi, Delhi-110032, India,*
[5] *Inha University in Tashkent, Ziyolilar 9, Tashkent 100170, Uzbekistan*
[6] *Central Asian University, Milliy Bog' Street 264, Tashkent 111221, Uzbekistan*
[7] *National University of Uzbekistan, Tashkent 100174, Uzbekistan*
[8] *Department of Physics, Zhejiang Normal University, Jinhua 321004, People's Republic of China*
[9] *New Uzbekistan University, Mustaqillik ave. 54, 100007 Tashkent, Uzbekistan*

arXiv:2308.13901v1 [gr-qc] 26 Aug 2023

The current manuscript deals with the tidal force effects, geodesic deviation, and shadow constraints of the Schwarzschild-like black hole theorised in Starobinsky-Bel-Robinson gravity exhibiting M-theory compactification. In the current analysis, we explore the radial and angular tidal force effects on a radially in-falling particle by the central black hole, which is located in this spacetime. We also numerically solve the geodesic deviation equation and study the variation of the geodesic separation vector with the radial coordinate for two nearby geodesics using suitable initial conditions. All the obtained results are tested for Sag A* and M87* by constraining the value of the stringy gravity parameter β using the shadow data from the event horizon telescope observations. All the results are compared with Schwarzschild black hole spacetime. In our study, we found that both the radial and angular tidal forces experienced by a particle switch their initial behaviour and turn compressive and stretching, respectively, before reaching the event horizon. The geodesic deviation shows an oscillating trend as well for the chosen initial condition. For the constrained value of β, we see that the spacetime geometry generated by Sag A* and M87* is effectively same for both Schwarzschild and Starobinsky-Bel-Robinson black hole. Furthermore, we also calculated the angular diameter of the shadow in Starobinsky-Bel-Robinson black hole and compared with the Schwarzschild black hole. It is observed that the angular diameter of shadow for M87* and Sgr A* in Starobinsky-Bel-Robinson black hole is smaller than the Schwarzschild black hole. The calculated results satisfy the event horizon telescope observational constraints. Finally, we have concluding remarks.

I. INTRODUCTION

Black holes are one of the most fascinating components in our universe, representing a fundamental prediction within Einstein's theory of general relativity. The scientific community and researchers have been intrigued by black holes for numerous years, driven by various reasons. Their captivating nature makes them curious and exotic entities within the universe. A widely accepted theory is that black holes are formed through the gravitational collapse of massive stars [1]. The existence of black holes has been strongly supported by several observational characteristics. Notably, measurements of black hole spin in X-ray binaries, the detection of gravitational wave signals from binary black hole mergers by LIGO [2–4], and the remarkable image of a black hole at the core of the galaxy M87 captured by the Event Horizon Telescope (EHT) collaboration [5] have provided compelling evidence. Additionally, the discovery of a wide star-black hole binary system through radial velocity measurements [6] has further contributed to our comprehensive understanding of black holes. These achievements collectively reinforce the evidence for the existence and properties of black holes. These enigmatic objects have captured considerable attention across the fields of astrophysics, astronomy and high-energy physics. Through studying black holes, researchers have made remarkable discoveries in areas such as thermodynamics, quantum effects, and gravitational interactions within curved spacetime. The physical properties of black holes remain a fascinating subject for the scientific community, even after the first discovery of a black hole as a solution to Einstein's equations more than a hundred years ago in General Relativity (GR). This initial discovery gave rise to the well-known Schwarzschild solution [7], characterized by a single parameter - the mass of the black hole. Since then, multiple solutions to Einstein's equations have been unveiled, going beyond the Schwarzschild solution. These encompass the Kerr black hole, exhibiting rotation [8], the Reissner-Nordström black hole, demonstrating electrical charge [9, 10], and the Kerr-Newman black

* arora09dhruv@gmail.com
† niyazuddin182@gmail.com
‡ himanshuch1729@gmail.com
§ ujjaldebnath@gmail.com
¶ atamurotov@yahoo.com
** gmustafa3828@gmail.com

hole, possessing both charge and rotation [11]. Thus in the context of GR, black holes can be described using four fundamental parameters: mass, charge, rotation, and angular momentum. However, when it comes to astrophysical black holes, it is generally anticipated that they are characterized by only two of these parameters - mass and angular momentum. Despite this simplification, these black holes may exist within highly dynamic environments, featuring elements like accretion disks and electromagnetic fields [12]. Furthermore, in alternative theories of gravity, black holes are linked to slight deviations from their standard counterparts, leading to unique features in their behavior [13]. Investigating how these deviations can be observed in electromagnetic and gravitational wave observations is currently a highly active area of research in the scientific literature [14–16]. In the last decade, modified gravity models have been extensively explored to address various cosmological issues and investigate the properties of black holes. Numerous theories have been proposed, taking diverse approaches and methods. Among these, Einstein Gauss-Bonnet (EGB) gravity, which finds support from string theory and related dual models, including M-theory [17–19], has emerged as one of the most extensively studied theory. In the context of EGB gravity, black holes have been investigated with respect to the Gauss-Bonnet gravity parameter. This parameter plays a pivotal role in constructing and studying black holes within these gravity models [20, 21]. The thermodynamics and optical properties of these black holes have been carefully examined concerning the Gauss-Bonnet parameter[22, 23]. Notably, it has been observed that variations in the Gauss-Bonnet parameter significantly impact the physical properties of black holes, with the shadow radius decreasing as the parameter increases [24]. These findings provide valuable insights into the interplay between gravity models and the intriguing behavior of black holes, furthering our understanding of the cosmos at both macroscopic and microscopic scales. Recently, a novel gravitational model called Starobinsky-Bel-Robinson (SBR) gravity has been introduced, incorporating a new parameter denoted by β [25]. This four-dimensional gravity is inspired by M-theory, which, at lower energies, is described by eleven-dimensional supergravity. Specifically, there are suggestions that this modified gravity can be derived from M-theory compactified on a two-sphere factor represented by $S^3 \times S^4$ [25]. Various models have been proposed within this novel gravitational framework. Inflation scenarios involving the β parameter have been studied, and associated cosmological observables have been computed [26]. These investigations have explored how observational data can impose constraints on the value of β. Furthermore, Schwarzschild-type black holes in SBR gravity have been constructed, and their thermodynamic properties have been analyzed in terms of the parameter β. Relevant quantities, including entropy and pressure, have been computed, revealing that the presence of β introduces corrections to these quantities [27]. These explorations shed light on the influence of the β parameter on the thermodynamic properties of black holes in the context of SBR gravity. In a recent study, Belhaj et al. [28] investigated the deflection angle and shadows produced by black holes within the framework of Starobinsky-Bel-Robinson Gravity derived from M-theory.

In the current paper, we focus on the analysis of tidal forces in the black holes in Starobinsky-Bel-Robinson modified Gravity from M-theory, recently obtained by Delgadoa and Ketov [27]. It is widely known that tidal forces within a Schwarzschild spacetime cause a body falling towards the event horizon to experience stretching in the radial direction while being compressed in the angular directions [29, 30]. The stretching and compression are induced by gravity's tidal effect, which is created by a difference in gravity's strength between two neighbouring places. This phenomenon is highly widespread in our cosmos and have sparked public scientific interest for most of the twentieth century. In the last decades, tidal forces were analyzed for Kiselev black hole [31], Hayward black hole spacetime [32], null naked singularity spacetime [33] and some regular black holes [34]. The authors of [35] compared the tidal force effects in a stringy charged black hole to those in Schwarzschild and RN black holes. The authors of [36] studied tidal forces in Reissner-Nordström spacetime and discovered that the radial and angular tidal forces switches sign at the event horizon. Also, the tidal forces in presence of a cosmological constant were studied in [37].

The tidal forces in the Kerr Black Hole were investigated in Refs.[38–43]. Tidal forces play a significant role in astrophysics, leading to intriguing phenomena such as tidal disruption events (TDEs). In these events, a star may be disrupted by the tidal forces exerted by a black hole[44] ,giving rise to bright flares of X-ray [45], ultraviolet [46], and optical [47] radiation. These TDEs offer valuable insights into the interactions between black holes and surrounding celestial bodies.

In this paper, we thoroughly analyze the tidal forces and geodesic deviation for a radially in-falling particle in the Schwarzschild-like black hole spacetime in the Starobinsky-Bel-Robinson Gravity. Furthermore, we also find the constraint of the black hole parameter β with the black hole shadow by the supermassive black holes $M87^*$ and $SgrA^*$ data. The shadow of a black hole is created by photons emitted from the central black hole and eventually reaching an observer. These photons follow unstable orbits, forming light rings around the black hole. When projected onto an observing screen, these light rings give rise to the shadow. Unstable orbits are critical because small perturbations can cause photons to either be captured by the central compact

object or escape to infinity. The light rays composing the shadow pass very close to the event horizon and experience significant bending due to the strong gravitational lensing effect. Consequently, observations of the shadow provide a means to test the strong-field properties of gravity. The seminal work of Synge [48] and Luminet [49] was instrumental in developing a formula for calculating the angular radius of photons. This formula, in turn, facilitated the characterization of the region surrounding a Schwarzschild black hole by elucidating the behavior of the light deflection angle as it diverges. Following this, Bardeen's[50] groundbreaking research in 1973 delved into the intricacies of the shadow cast by a spinning Kerr black hole, demonstrating how the black hole's rotation caused distortions in the shadow's shape. Within this context, the photon ring, encircling the black hole's shadow, emerges as a pivotal element that can be harnessed to determine the black hole's parameters and unveil gravitational features within the information field. Subsequently, there has been a surge of activity in both analytical/numerical investigations and observational studies related to the shadows cast by the different kind of black holes [24, 51–56]. Numerous studies have been conducted in this area, including those by the works of literature [57–63].

The organization of this paper is as follows. In section II, we reconsider the nature of the black holes in the SBR gravity. In Section III we discuss the radial and angular tidal forces in SBR black hole spacetime using tetrad formalism. Section IV presents the numerical solution for the geodesic deviation equation and compares the results with Schwarzschild spacetime. In section V, we investigate the shadow behaviors and provide predictions for β by the help of EHT data. We conclude our results in section VI. Throughout the manuscript we use geometrized units as $G = c = 1$.

II. BLACK HOLES IN SBR GRAVITY

In this section, we discuss about the Schwarzschild like black hole solution in SBR gravity which is theorized to be embedded in M-theory existing in eleven dimensional spacetime. This solution is discussed by the authors in [64]. M-theory has a bosonic sector having a metric g_{MN} and a tensor 3 -form C_{MNP} coupled to M2-branes which are dual to M5-branes [65]. Using the compactification process coupled with the existence of stringy fluxes, the corresponding 4-D gravity models may be produced [26, 27, 64]. The action considered is given by:

$$S_{SBR} = \frac{M_{pl}}{2}\int d^4x\sqrt{-g}\left(R + \frac{R^2}{6m^2} - \frac{\beta}{32m^6}\left(P_4^2 - E_4^2\right)\right). \tag{1}$$

where g is the metric determinant and R is the Ricci curvature scalar. m is a free mass parameter which could have various interpretations depending on the underlying theory. β is a positive dimensionless coupling whose value is determined by the compactification of M-theory, and could be fixed by studying the optical behaviors of black hole. P_4^2 and E_4^2 are the Pontryagin and the Euler topological densities which are related to the Bel-Robinson tensor $T_{\mu\nu\lambda\rho}$ in four dimensions by means of the relation [26, 27, 64]

$$T^{\mu\nu\delta\rho}T_{\mu\nu\delta\rho} = \frac{1}{4}\left(P_4^2 - E_4^2\right). \tag{2}$$

The Bel-Robinson tensor is defined as:

$$T^{\mu\nu\delta\sigma} = R^{\mu\rho\gamma\delta}R^{\nu\sigma}_{\rho\gamma} + R^{\mu\rho\gamma\sigma}R^{\nu\delta}_{\rho\gamma} - \frac{1}{2}g^{\mu\nu}R^{\rho\gamma\alpha\delta}R^{\sigma}_{\rho\gamma\alpha}. \tag{3}$$

It can be seen from Eq.(1) that the gravity action depends on two parameters (m, β). This gives rise to plethora of applications such as studying Hawking radiation [66], entropy [67] and inflation [68]. The black hole solution in SBR gravity depends on the value of β corrected to first order perturbations. The line element of this non-rotating solution has been found to be [27]

$$ds^2 = -f(r)dt^2 + \frac{1}{f(r)}dr^2 + r^2d\Omega^2. \tag{4}$$

where the metric component $f(r)$ is given by

$$f(r) = 1 - \frac{2M}{r} + \beta\left(\frac{8\sqrt{2}\pi M}{r^3}\right)^3\left(\frac{108r - 194M}{5r}\right). \tag{5}$$

where $r_s = 2M$ is the Schwarzschild radius. Here, M represent the mass parameter.

III. TIDAL FORCE EFFECTS IN SBR BLACK HOLES

In this section, we will delve into the complexities of the equations involving tidal forces in our spacetime. To investigate the equation for the distance between two infinitesimally close and free falling particles, we use the following equation for the spacelike components of the geodesic deviation vector η^μ

$$\frac{D^2\eta^\mu}{D\tau^2} = R^\mu_{\alpha\beta\gamma}v^\alpha v^\beta\eta^\gamma, \tag{6}$$

where $R^\mu_{\alpha\beta\gamma}$ is the Riemann curvature tensor and v^μ is the unit tangent vector to the geodesic. It is known that a test body moving in space solely under the influence of gravity follows a geodesic [69]. Now, due to the difference in curvature at every point on the geodesic and the fact that each point on the test body tends to follows a unique geodesic, the body experiences a difference in acceleration, leading to a stretching and

squeezing effect known as tidal forces. This is the reason that the Riemann curvature tensor is used to study the tidal interactions due to gravity.

To calculate the tidal forces in the falling body's frame, we make use of the tetrad formalism [70]. Tetrads are geometric objects that form a set of local coordinate bases, i.e. a locally defined set of four linearly independent vector fields known as tetrads or vierbien. At each point on a geodesic, there is a tetrad frame that forms a local inertial reference frame, where the laws of special relativity apply. Since, Lorentz transformations can connect an infinite number of orthonormal bases at a specific point, they cannot give enough information about the connection on the metric. To resolve this, we use the transformation from the orthonormal basis to the coordinate basis:

$$\overrightarrow{e}_{\mu} = \hat{e}^{\hat{i}}_{\mu}\overrightarrow{e}_{\hat{i}}, \tag{7}$$

where $\overrightarrow{e}_{\mu}$ represents the coordinate basis, $\overrightarrow{e}_{\hat{i}}$ represents the orthonormal basis and $\hat{e}^{\hat{i}}_{\mu}$ are the tetrad components. The metric tensor components in the tetrad basis are given as [70]:

$$g_{\mu\nu} = \eta_{\hat{i}\hat{j}}\hat{e}^{\hat{i}}_{\mu}\hat{e}^{\hat{j}}_{\nu}. \tag{8}$$

This is the key equation that enables us to use orthonormal basis in curved spacetime. The tetrad components for freely falling frames can be obtained from Eq. (7) and Eq. (8).

$$\hat{e}^{\mu}_{\hat{0}} = \left\{\frac{E}{f(r)}, -\sqrt{E^2 - f(r)}, 0, 0\right\}, \tag{9}$$

$$\hat{e}^{\mu}_{\hat{1}} = \left\{\frac{-\sqrt{E^2 - f(r)}}{f(r)}, E, 0, 0\right\}, \tag{10}$$

$$\hat{e}^{\mu}_{\hat{2}} = \left\{0, 0, \frac{1}{r}, 0\right\}, \tag{11}$$

$$\hat{e}^{\mu}_{\hat{3}} = \left\{0, 0, 0, \frac{1}{r\sin\theta}\right\}, \tag{12}$$

where the tetrad components follow the following rule:

$$\hat{e}^{\mu}_{\hat{i}}\hat{e}^{\hat{j}}_{\mu} = \delta^{\hat{j}}_{\hat{i}}. \tag{13}$$

We notice that the component $\hat{e}^{\mu}_{\hat{0}} = v^{\mu}$ is the tangent vector to the geodesic i.e. the four velocity of the particle. Also, the geodesic deviation vector follows the transformation rule from global coordinates to local orthonormal coordinates:

$$\eta^{\mu} = \hat{e}^{\mu}_{\hat{i}}\eta^{\hat{i}}. \tag{14}$$

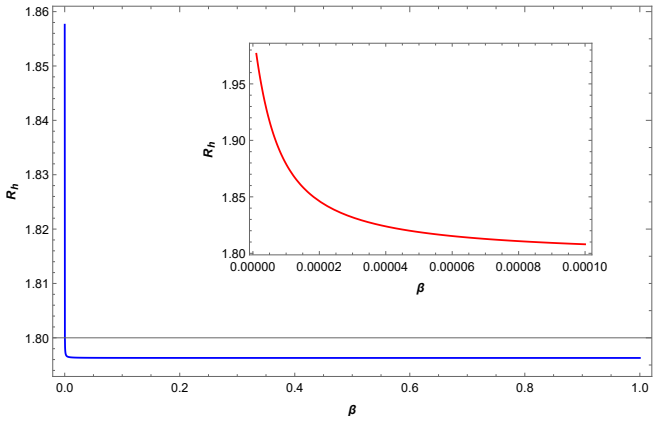


FIG. 1. Event horizon location in SBR black holes for different values of stringy parameter β. We take M=1 for computation. It is noticed that the event horizon begins shifting for values of the order $\sim 10^{-4}$

The non vanishing independent components of Riemann tensor for spherically symmetric spacetimes, including SBR black hole spacetime are given by [71]:

$$R^{1}_{212} = \frac{-rf'(r)}{2}, \tag{15}$$

$$R^{1}_{010} = \frac{f(r)f''(r)}{2}, \tag{16}$$

$$R^{1}_{313} = \frac{-rf'(r)\sin^2\theta}{2}, \tag{17}$$

$$R^{2}_{020} = R^{3}_{030} = \frac{f(r)f'(r)}{2r}, \tag{18}$$

$$R^{2}_{323} = \sin^2\theta\left(1 - r^2 f(r)\right) \tag{19}$$

For computing the Riemann curvature tensor components in tetrad basis, we use the tetrad formalism as:

$$R^{\hat{i}}_{\hat{j}\hat{k}\hat{l}} = R^{\mu}_{\alpha\beta\gamma}e^{\hat{i}}_{\mu}e^{\alpha}_{\hat{j}}e^{\beta}_{\hat{k}}e^{\gamma}_{\hat{l}}. \tag{20}$$

Following the Eq. (20), the tidal tensor components are given by:

$$R^{\hat{1}}_{\hat{0}\hat{1}\hat{0}} = \frac{f''(r)}{2}, \tag{21}$$

$$R^{\hat{2}}_{\hat{0}\hat{2}\hat{0}} = R^{\hat{3}}_{\hat{0}\hat{3}\hat{0}} = \frac{f'(r)}{2r}, \tag{22}$$

A. TIDAL FORCE EQUATIONS

Upon obtaining the expressions from Eq. (21-22), we can obtain the relative acceleration between two nearby particles as:

$$\frac{D^2\eta^{\hat{r}}}{D\tau^2} = \frac{-f''(r)}{2}\eta^{\hat{r}}, \tag{23}$$

$$\frac{D^2\eta^{\hat{\theta}}}{D\tau^2} = -\frac{f'(r)}{2r}\eta^{\hat{\theta}}, \tag{24}$$

$$\frac{D^2\eta^{\hat{\phi}}}{D\tau^2} = -\frac{f'(r)}{2r}\eta^{\hat{\phi}}. \tag{25}$$

By substituting the values of $f(r)$ and its higher derivatives in Eq. (23-25) we get:

$$\frac{D^2\eta^{\hat{r}}}{D\tau^2} = \frac{4Mr^9 - 243\cdot 2^{\frac{27}{2}}\pi^3M^3\beta r + 1067\cdot 2^{\frac{25}{2}}\pi^3M^4\beta}{2r^{12}}\eta^{\hat{r}} \tag{26}$$

$$\frac{D^2\eta^{\hat{\theta}}}{D\tau^2} = \frac{-10Mr^9 + 243\cdot 2^{\frac{25}{2}}\pi^3M^3\beta r - 485\cdot 2^{\frac{25}{2}}\pi^3M^4\beta}{10r^{12}}\eta^{\hat{\theta}} \tag{27}$$

$$\frac{D^2\eta^{\hat{\phi}}}{D\tau^2} = \frac{-10Mr^9 + 243\cdot 2^{\frac{25}{2}}\pi^3M^3\beta r - 485\cdot 2^{\frac{25}{2}}\pi^3M^4\beta}{10r^{12}}\eta^{\hat{\phi}} \tag{28}$$

B. RADIAL TIDAL FORCE

Following the expression in Eq.(23) and we notice that the radial tidal force vanishes at a point $R = R_0^{rtf}$ given by solving the below equation:

$$R_0^{rtf} = \frac{4M}{r^3} - f''(\beta, r) = 0. \tag{29}$$

where the function $f(\beta, r)$ is taken for simplicity in calculation and representation, and is given by:

$$f(\beta, r) = \beta\left(\frac{8\sqrt{2}\pi M}{r^3}\right)^3\left(\frac{108r - 194M}{5r}\right). \tag{30}$$

It can be seen from the graph in fig.(2) that as the value of β is increased, the peak value of tidal force shifts to the right. This suggests that the maximum radial stretching happens far from the horizon as the departure from Schwarzschild geometry is significant. The peak value of radial tidal force can be calculated by equating the first derivative of the term in RHS of Eq.(23) to zero.

$$R_{peak}^{rtf} = \frac{d}{dr}\left(\frac{\frac{4M}{r^3} - f''(\beta, r)}{2}\right) = 0. \tag{31}$$

The peculiar trend that can be noticed from the graph is that the nature of force changes from radial stretching to compression for non zero values of β. This is in strict contrast to the Schwarzschild black hole case where the radial tidal force profile shows a monotonically increasing trend and becomes infinity as $r \to 0$. The transition from radial stretching to compression reverts back very close to the event horizon, after which the tidal force increases drastically till infinity.

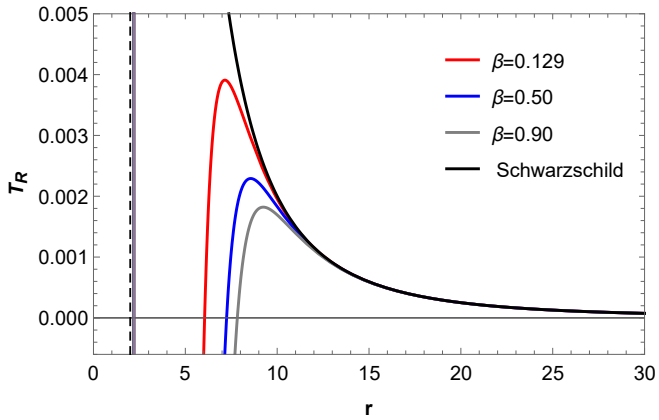


FIG. 2. Radial tidal force as a function of the radial coordinate. We notice that the force vanishes at a single point in contrast to what happens with Schwarzschild spacetime. We take $M = 1$ and $\beta = 0.20$. Dotted line shows the location of event horizon for Schwarzschild spacetime

We test the behavior of radial tidal force for both Sag A* and M87* and the results are displayed in figures 3 and 4 respectively.

For both of these cases, the force profile for SBR black hole coincides perfectly with the Schwarzschild black hole case in the range $-0.00038 < \beta < 0.00089$ and $-0.002 < \beta < 0.0003$ for Sag A* and M87* respectively. This suggests that the spacetime geometry generated by Sag A* and M87* is best explained by the Schwarzschild geometry.

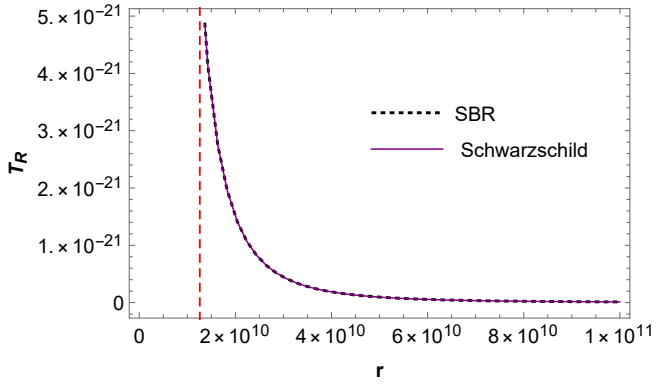


FIG. 3. Radial tidal force profile for Sag A* as a function of the radial coordinate. Dotted red line shows the location of event horizon for Schwarzschild spacetime

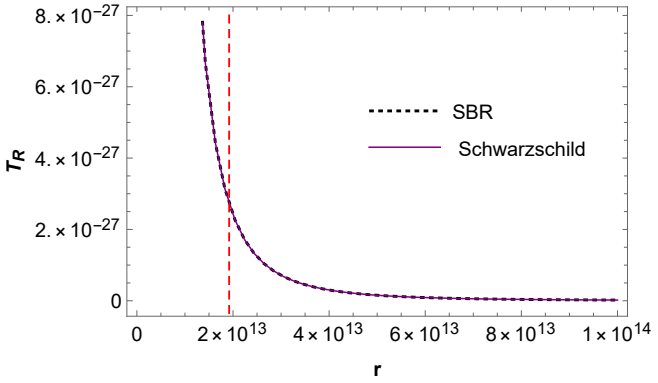


FIG. 4. Radial tidal force profile for M87* as a function of the radial coordinate. Dotted red line shows the location of event horizon for Schwarzschild spacetime

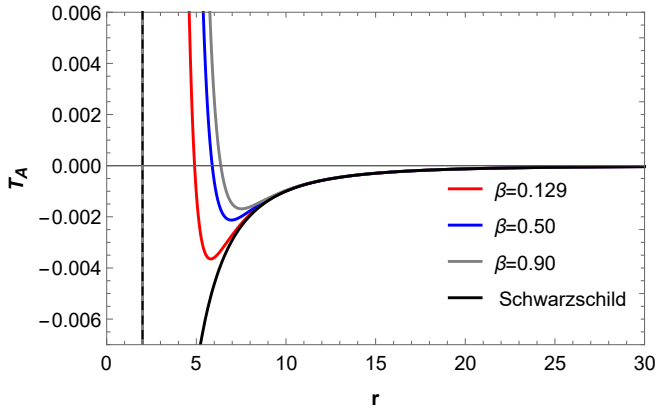


FIG. 5. Angular tidal force as a function of the radial coordinate. We notice that the force vanishes at a single point in contrast to what happens with Schwarzschild spacetime. We take $M = 1$ and $\beta = 0.20$. Dotted line shows the location of event horizon for Schwarzschild spacetime

C. ANGULAR TIDAL FORCE

Just similar to the radial tidal force profile, the angular force profile also vanishes at a point $R = R_0^{atf}$. It can be computed by equating the RHS of Eq.(24) to zero and is given by:

$$R_0^{atf} = \frac{2M}{r^2} + f'(\beta, r) = 0. \tag{32}$$

For simplicity, the term $f'(r)$ is expressed as $\frac{2M}{r^2} + f'(\beta, r)$. where $f(\beta, r)$ is given by the Eq.(30). Angular tidal forces show compressive behavior as the particle falls radially inwards from infinity. The force reaches a peak value and inverts it nature to stretching. The peak value shifts to the right for increasing value of β similar to the case of radial force profile. This suggests that maximum compression happens far from the horizon as the departure from the Schwarzschild geometry increases.

Following the convention in Eq.(30), the peak value of angular tidal force can be calculated by equating the first derivative of the expression in Eq.(24) to zero.

$$R_{peak}^{atf} = \frac{d}{dr}\left(\frac{\frac{-2M}{r^2} - f'(\beta, r)}{2r}\right) = 0. \tag{33}$$

Angular force profile also shows a very different trend from the Schwarzschild geometry. In Schwarzschild black hole case, the angular tidal force becomes increasingly compressive as the radial coordinate becomes smaller only to reach infinity as $r \to 0$. But, in the case of SBR black hole, the angular tidal force switches its nature from compressive to stretching as the radial coordinate decreases for in-falling particle. The graph in fig.(5) shows the general force profile. The compression to stretching trend again reverts back very close to the horizon, tending to infinity thereafter. We test the behavior of angular tidal force for both Sag A* and M87* and the results are displayed in figures 6 and 7 respectively. We find that for both the cases, the SBR and Schwarzschild force profile coincides. This confirms that the spacetime geometry generated by both compact objects resembles Schwarzschild geometry.

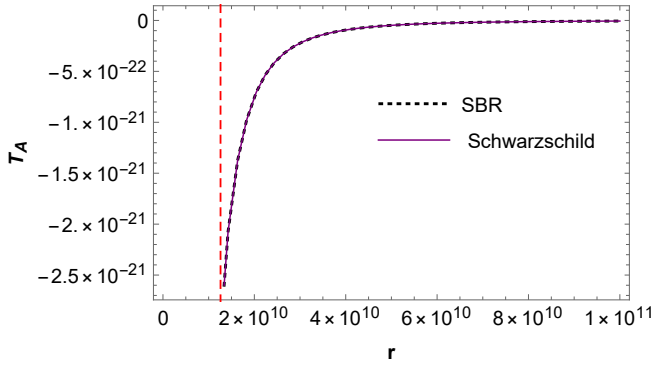


FIG. 6. Angular tidal force profile for Sag A* as a function of the radial coordinate. Dotted red line shows the location of event horizon for Schwarzschild spacetime

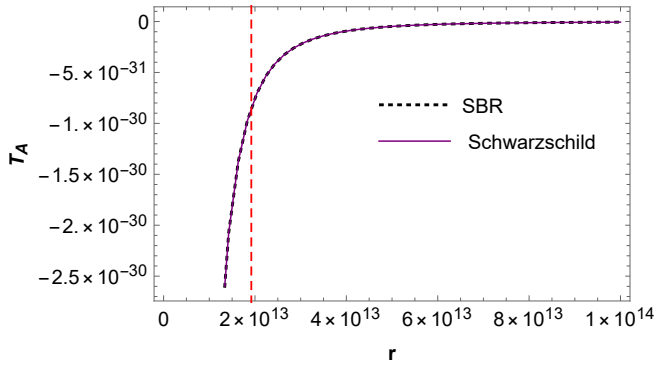


FIG. 7. Angular tidal force profile for M87* as a function of the radial coordinate. Dotted red line shows the location of event horizon for Schwarzschild spacetime

IV. GEODESIC DEVIATION

In this section we discuss the variation of geodesic deviation vector with the radial coordinate in SBR black hole spacetime. The deviation vector measures the deformation of a body falling radially in any spacetime geometry. We can convert Eq. (23) and (24) in second derivatives w.r.t. r by substituting $dr/d\tau = -\sqrt{E^2 - f(r)}$ which results from Eq. (9). This gives us the following second order differential equations in r:

$$(E^2 - f(r))\frac{D^2\eta^{\hat{r}}}{Dr^2} - \frac{f'(r)}{2}\frac{D\eta^{\hat{r}}}{dr} + \frac{f''(r)}{2}\eta^{\hat{r}} = 0, \quad (34)$$

$$(E^2 - f(r))\frac{D^2\eta^{\hat{i}}}{Dr^2} - \frac{f'(r)}{2}\frac{D\eta^{\hat{i}}}{dr} + \frac{f'(r)}{2r}\eta^{\hat{i}} = 0. \quad (35)$$

where $i = \{\theta, \phi\}$. The analytic solution for Eq. (34) and (35) as pointed out in [72] can be given as:

$$\eta^{\hat{r}}(r) = \sqrt{E^2 - f(r)}\left[C_1 + C_2 \int \frac{dr}{(E^2 - f(r))^{3/2}}\right], \quad (36)$$

$$\eta^{\hat{i}}(r) = \left[C_3 + C_4 \int \frac{dr}{r^2\sqrt{(E^2 - f(r))}}\right] r. \quad (37)$$

where C_1, C_2, C_3, C_4 are constants of integration. In order to find their value, we numerically solve the differential equations by imposing some initial conditions. For the purpose of this study we take the following initial conditions.

$$\eta^{\hat{\beta}}(b) > 0, \dot{\eta}^{\hat{\beta}}(b) = 0. \quad (38)$$

where $\beta = \{r, \theta, \phi\}$. $\eta^{\hat{\beta}}(b)$ represents the separation between two nearby geodesics at $r = b$ in the radial and angular directions. The initial condition in Eq. (38) represents a mass released from rest at $r = b$. In the next sections we will discuss the components of radial and the angular deviation vectors in detail.

A. RADIAL COMPONENT

In Fig.(8) we have shown the radial component of the geodesic deviation vector after solving Eq. (36) with initial condition. We notice that for non zero values for β, the radial component of the geodesic deviation vector reaches a finite value outside the event horizon, falls rapidly and again peaks just inside the horizon as $r \to 2M$, unlike in the Schwarzschild black hole scenario, where it crosses the event horizon and goes till infinity because of the infinite stretching radial tidal force at the singularity. This peak value attained outside the horizon keeps on decreasing as the value of β is increased.

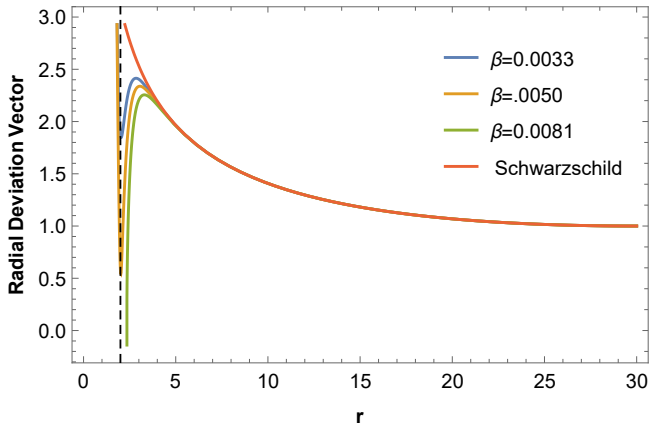


FIG. 8. Radial geodesic deviation profile for a radially in-falling particle in SBR black hole spacetime. We use M=1 to show the trend for different values of β. Dotted line shows the location of event horizon in Schwarzschild spacetime.

By using the constrained value of β, we show the radial geodesic deviation trend for Sag A* and M87* in Fig (9) and Fig.(10). As seen from figures, the SBR black hole and Schwarzschild black hole lines perfectly coincide, leading to the conclusion that Schwarzschild geometry perfectly defines the geodesic deviation for a test particle falling radially in the spacetime generated by Sag A* and M87* for $-0.00038 < \beta < 0.00089$ and $-0.002 < \beta < 0.0003$ respectively.

B. ANGULAR COMPONENT

In Fig.(11) we have shown the angular component of geodesic deviation vector by constraining the Eq. (37) with the initial condition. The results are compared with the Schwarzschild black hole. It is observed that in the SBR black hole case, the angular deviation decreases for a radially in-falling particle, shows a little spike near the event horizon and then continues to fall beyond the event horizon. This is in contrast to the Schwarzschild black hole scenario in which the deviation decreases throughout the range of the radial coordinate. The graph in Fig.(11) shows the behavior of the geodesics. The peak

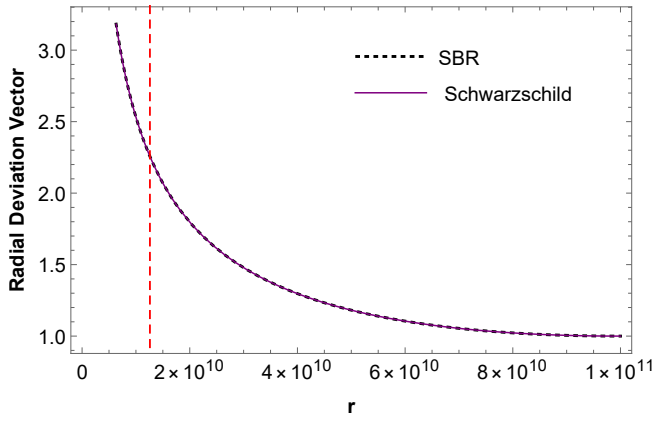


FIG. 9. Radial geodesic deviation profile for Sag A* using the constrained values of β. The dotted red line shows the location of the event horizon of Schwarzschild spacetime.

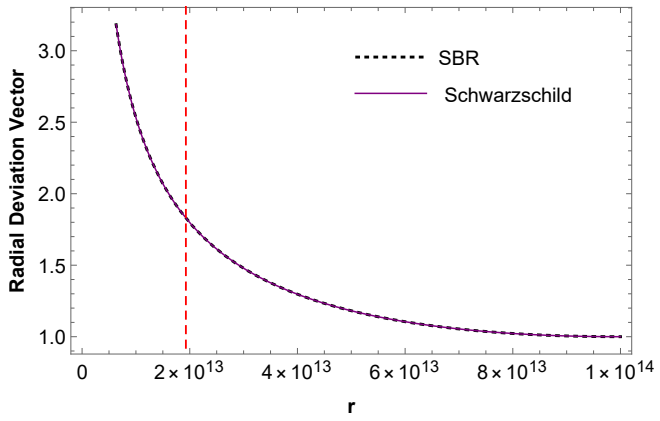


FIG. 10. Radial geodesic deviation profile for M87* using the constrained values of β. The dotted red line shows the location of the event horizon of Schwarzschild spacetime.

value is higher for increasing value of β. By using the constrained value of β, we show the angular geodesic deviation trend for Sag A* and M87* in Fig (12) and Fig.(13). As seen from the graphs below, the SBR black hole and Schwarzschild black hole lines perfectly coincide, leading to the conclusion that Schwarzschild geometry perfectly defines the geodesic deviation for a test particle falling radially in the spacetime generated by Sag A* and M87* for $-0.00038 < \beta < 0.00089$ and $-0.002 < \beta < 0.0003$ respectively.

V. BLACK HOLE SHADOW WITH M87* AND SGR A*

In this section we explore the shadow of the BH SBR gravity from M-theory. We know that for the angular radius of the BH shadow we study [57, 73]

$$\sin^2\alpha_{sh} = \frac{h(r_{ph})^2}{h(r_{obs})^2}, \tag{39}$$

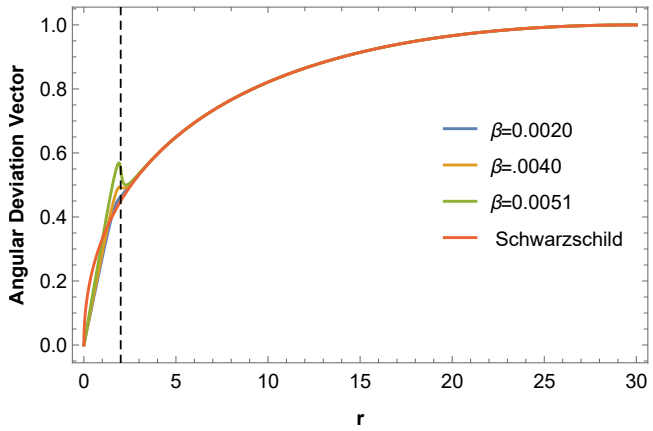


FIG. 11. Angular geodesic deviation profile for a radially infalling particle in SBR black hole spacetime. We use M=1 to show the trend for different values of β. Dotted line shows the location of event horizon in Schwarzschild spacetime.

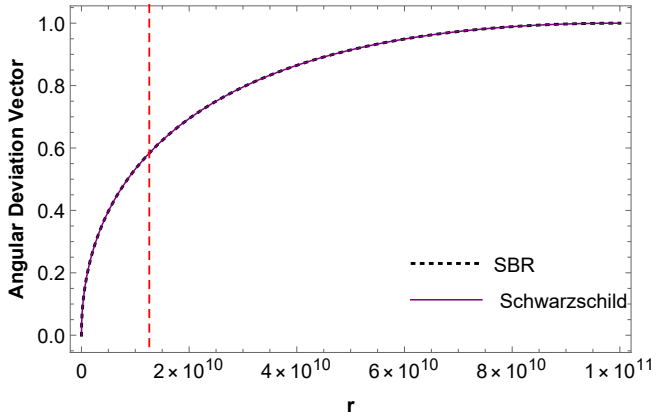


FIG. 12. Angular geodesic deviation profile for Sag A* using the constrained values of β. The dotted red line shows the location of the event horizon of Schwarzschild spacetime.

with

$$h(r)^2 = \frac{r^2}{f(r)}, \tag{40}$$

where r_{ph}, r_{obs} are the radius of photon sphere and observer distance, respectively. $h(r)$ was early introduced in several references (for example, see [73]). α_{sh} is the angular radius of the BH shadow.

Here we combine Eqs. (39) and (40), and for a distant observer the Eq. (39) takes the following form

$$\sin^2\alpha_{sh} = \frac{r_{ph}^2}{f(r_{ph})}\frac{f(r_{obs})}{r_{obs}^2}. \tag{41}$$

One can easily find the observable radius of BH shadow for observer at infinite distance using Eq. (41) in the following form as[57]

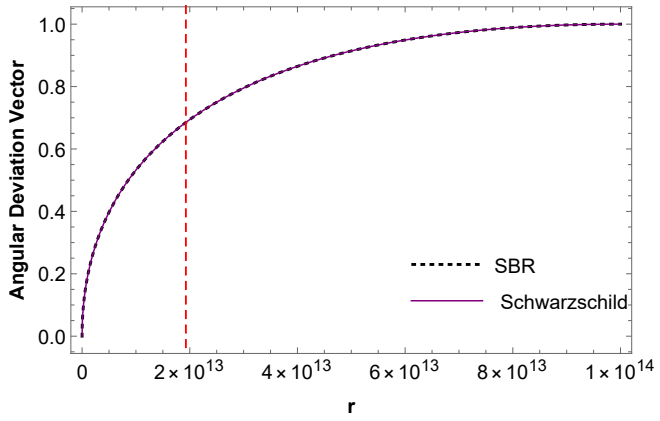


FIG. 13. Angular geodesic deviation profile for M87* using the constrained values of β. The dotted red line shows the location of the event horizon of Schwarzschild spacetime.

$$R_{sh} \simeq r_{obs}\sin\alpha_{sh} \simeq \frac{r_{ph}}{\sqrt{f(r_{ph})}}. \tag{42}$$

Now we consider that the supermassive BHs M87* and Sgr A* are spherically symmetric static and SBR parameter β from M-theory. Although the observation got by the EHT collaboration does not support the assumption taken here. However, here we explore theoretically the constrain on the parameter β, from the data provided by the EHT project. To constrain this parameter β in SBR gravity we use the observational data released by the EHT project for the BH shadows of the supermassive BHs M87* and Sgr A*. The angular diameter of the shadow, the distance from sun system and the mass of of the BH at the centre of the galaxy M87, are $\Omega_{\rm M87^*} = 42 \pm 3\ \mu$as, $D = 16.8 \pm 0.8$ Mpc and $M_{\rm M87^*} = (6.5 \pm 0.7)x10^9\ M_\odot$, respectively [5]. For the Sgr A* the data recently obtained by the EHT project is $\Omega_{\rm Sgr\ A^*} = 51.8 \pm 2.3\ \mu$as, $D = 8277 \pm 9 \pm 33$ pc and $M_{\rm Sgr\ A^*} = 4.297 \pm 0.013x10^6\ M_\odot$ (VLTI) [74]. Using this data, we can estimate the diameter of the shadow cast by the BH, per unit mass from the following expression [24, 75],

$$d_{\rm sh} = \frac{D\Omega}{M}\,. \tag{43}$$

Now we can obtain the diameter of the shadow from the expression $d_{\rm sh}^{\rm theo} = 2R_{\rm sh}$. Thus, the diameter of the BH shadow image is $d_{\rm sh}^{\rm M87^*} = (11 \pm 1.5)M$ for M87* and $d_{\rm sh}^{\rm Sgr\ A^*} = (9.5 \pm 1.4)M$ for Sgr A*. From the data by the EHT collaboration, we obtain the constrain on the parameter β for the supermassive BHs at the centre of the galaxy M87* and the Sgr A*. We present our results obtained here in the Fig. 14. In this figure,it is observed that the angular diameter of BH shadow decreases with the increasing value of parameter β. It is also observed that the angular diameter of shadow for M87* and

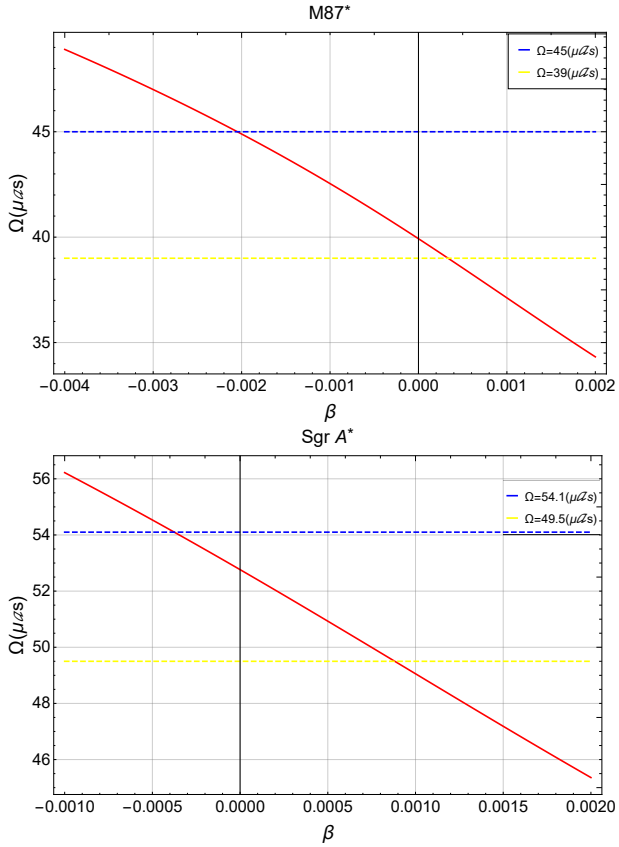


FIG. 14. The angular diameter of shadow for SBR black hole as a function of parameter β where angular diameter $\Omega = 42 \pm 3\mu as$ for M87* (upper panel); and for $\Omega = 51.8 \pm 2.3\mu as$ for Sgr A* (lower panel).

Sgr A* BH in the context of the SBR black hole BH is smaller than the other ordinary astrophysical BH such as Schwarzschild BH. Consequently, the SBR black hole parameter β has been constrained as $-0.00038 < \beta < 0.00089$ for Sgr A* and $-0.002 < \beta < 0.0003$ for M87* respectively.

VI. CONCLUSIONS

In this work, we have discussed tidal force effects and constrained the shadow of Schwarzschild type black holes in Starobinsky-Bel-Robinson modified theory of gravity. This work has been done in an attempt to define some observational signatures that can be used to distinguish SBR gravity from Einstein's general theory of relativity. The metric for spherically symmetric and static spacetime in this gravity differs from the Schwarzschild metric by introducing a stringy gravity parameter β which has a positive value.

We see that both the radial and angular tidal force profiles depend on the value of β and approach infinity as $r \to 0$. Both force profiles show a varying trend

where the radial force becomes compressive and angular tidal force becomes of stretching nature as a particle radially falls towards the black hole. This is in contrast with Schwarzschild black hole case in general relativity, where the radial tidal force monotonically increases till infinity showing stretching and the angular tidal force monotonically increases till infinity and shows compression. Both tidal forces vanish at two points outside the event horizon for different values of β giving us an opportunity to observe this phenomenon in some cases. In case of SBR gravity, the event horizon of black hole forms prior to reaching a radial distance of $r = 2M$. It can be clearly seen from Fig.(1) that the location of event horizon does not change for larger values of β i.e. of the order of $\beta > 10^{-4}$. Another important observation that can be made from the radial and angular force profiles is that the the peak value is attained farther away when compared with the Schwarzschild black hole case.

In addition to this, we also examined the geodesic deviation for a radially in-falling particle in the Schwarzschild like spacetime in SBR gravity. We consider the initial condition where a body is released from rest at a far away point where the spacetime is Minkowskian. It is observed that the radial deviation vector acts very similar to that of the Schwarzschild black hole case i.e. increasing as the particle approaches the event horizon, but shows an oscillating trend outside the event horizon. Opposite behaviour can be seen for angular deviation. It mimics the Schwarzschild black hole as the particle approaches the event horizon, but shows a sudden peak just around it and continues to rapidly fall to zero. Also, we have studied the tidal forces and geodesic deviation for Sag A* and M87* in SBR gravity by constraining the value of β using shadow size from EHT observations. We can see that in the particular β range, the tidal force and geodesic deviation profiles and both Sag A* and M87* in SBR gravity perfectly co-incide with the Schwarzschild black hole case.

Furthermore, we also calculated the angular diameter of the shadow in SBR black hole and compared it to Schwarzschild BH. It is observed that the angular diameter of shadow for M87* and Sgr A* BH in SBR black hole is smaller than the Schwarzschild BH. Consequently, the SBR black hole parameter β has been constrained as $-0.00038 < \beta < 0.00089$ for Sgr A* and $-0.002 < \beta < 0.0003$ for M87* respectively.It suggests that such black holes satisfy the EHT observational constraints, and it may be possible to detect the SBR black hole and distinguish it from the other astrophysical BHs in the future.

ACKNOWLEDGEMENTS

N.U.M would like to thank CSIR, Govt. of India for providing Senior Research Fellowship (No. 08/003(0141))/2020-EMR-I). This research is partly supported by Research Grants FZ-20200929344 and F-FA-2021-510 of the Uzbekistan Ministry for Innovative Development.

[1] S. Chandrasekhar, *The mathematical theory of black holes*, Vol. 69 (Oxford university press, 1998).
[2] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].
[3] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. Lett. **125**, 101102 (2020), arXiv:2009.01075 [gr-qc].
[4] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **102**, 043015 (2020), arXiv:2004.08342 [astro-ph.HE].
[5] K. Akiyama and et al., ApJ. **875**, L1 (2019), arXiv:1906.11238 [astro-ph.GA].
[6] J. Liu *et al.*, Nature **575**, 618 (2019), arXiv:1911.11989 [astro-ph.SR].
[7] K. Schwarzschild, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1916**, 189 (1916), arXiv:physics/9905030.
[8] R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[9] H. Reissner, Ann. Phys **50**, 106 (1916).
[10] G. Nordström, Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences **20**, 1238 (1918).
[11] E. T. Newman, R. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, J. Math. Phys. **6**, 918 (1965).
[12] J. Frank, A. King, and D. Raine, *Accretion power in astrophysics* (Cambridge university press, 2002).
[13] E. Berti *et al.*, Class. Quant. Grav. **32**, 243001 (2015), arXiv:1501.07274 [gr-qc].
[14] E. Barausse, V. Cardoso, and P. Pani, Phys. Rev. D **89**, 104059 (2014), arXiv:1404.7149 [gr-qc].
[15] E. Berti, R. Brito, C. F. B. Macedo, G. Raposo, and J. L. Rosa, Phys. Rev. D **99**, 104039 (2019), arXiv:1904.03131 [gr-qc].
[16] M. C. Ferreira, C. F. B. Macedo, and V. Cardoso, Phys. Rev. D **96**, 083017 (2017), arXiv:1710.00830 [gr-qc].
[17] J. T. Wheeler, Nucl. Phys. B **268**, 737 (1986).
[18] S. G. Ghosh and R. Kumar, Classical and Quantum Gravity **37**, 245008 (2020).
[19] S. G. Ghosh and S. D. Maharaj, Phys. Dark Univ. **30**, 100687 (2020), arXiv:2003.09841 [gr-qc].
[20] S. G. Ghosh, D. V. Singh, R. Kumar, and S. D. Maharaj, Annals Phys. **424**, 168347 (2021), arXiv:2006.00594 [gr-qc].
[21] D. V. Singh, B. K. Singh, and S. Upadhyay, Annals of Physics **434**, 168642 (2021).
[22] A. Belhaj and Y. Sekhmani, Annals of Physics **441**, 168863 (2022).
[23] A. Belhaj and Y. Sekhmani, Eur. Phys. J. Plus **137**, 278 (2022).
[24] S. Vagnozzi *et al.*, Class. Quant. Grav. **40**, 165007 (2023), arXiv:2205.07787 [gr-qc].

[25] S. V. Ketov, Universe **8**, 351 (2022), arXiv:2205.13172 [gr-qc].
[26] S. V. Ketov, E. O. Pozdeeva, and S. Y. Vernov, Journal of Cosmology and Astroparticle Physics **2022** (12), 032.
[27] R. C. Delgado and S. V. Ketov, arXiv preprint arXiv:2209.01574 (2022).
[28] A. Belhaj, H. Belmahi, M. Benali, Y. Hassouni, and M. Sedra, arXiv preprint arXiv:2304.03883 (2023).
[29] R. d'INVERNO, *Introducing Einstein's Relativity* (Clarendon Press,Oxford, 1992).
[30] G. E. M.P. Hobson and A. Lasenby, *General relativity: an introduction for physicists* (Cambridge University Press, New York, 2006).
[31] M. U. Shahzad and A. Jawad, Eur. Phys. J. C **77**, 372 (2017), arXiv:1706.00281 [gr-qc].
[32] H. C. Lima and L. C. Crispino, International Journal of Modern Physics D **29**, 2041014 (2020).
[33] S. Madan and P. Bambhaniya, arXiv preprint arXiv:2201.13163 (2022).
[34] M. Sharif and S. Sadiq, Journal of Experimental and Theoretical Physics **126**, 194 (2018).
[35] R. M. Gad, arXiv preprint arXiv:0708.2841 (2007).
[36] L. C. B. Crispino, A. Higuchi, L. A. Oliveira, and E. S. de Oliveira, Eur. Phys. J. C **76**, 168 (2016), arXiv:1602.07232 [gr-qc].
[37] V. Vandeev and A. Semenova, The European Physical Journal C **81**, 1 (2021).
[38] J.-A. Marck, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **385**, 431 (1983).
[39] D. Tsoubelis and A. Economou, General relativity and gravitation **20**, 37 (1988).
[40] C. Chicone, B. Mashhoon, and B. Punsly, Int. J. Mod. Phys. D **13**, 945 (2004), arXiv:astro-ph/0308421.
[41] C. Chicone and B. Mashhoon, Class. Quant. Grav. **22**, 195 (2005), arXiv:gr-qc/0409017.
[42] C. Chicone and B. Mashhoon, Annalen Phys. **14**, 290 (2005), arXiv:astro-ph/0404170.
[43] H. C. D. Lima Junior, L. C. B. Crispino, and A. Higuchi, Eur. Phys. J. Plus **135**, 334 (2020), arXiv:2003.09506 [gr-qc].
[44] M. Kesden, Phys. Rev. D **85**, 024037 (2012), arXiv:1109.6329 [astro-ph.CO].
[45] N. Bade, S. Komossa, and M. Dahlem, Astronomy and Astrophysics **309**, L35 (1996).
[46] S. Gezari, S. Basa, D. Martin, G. Bazin, K. Forster, B. Milliard, J. Halpern, P. Friedman, P. Morrissey, S. Neff, *et al.*, The Astrophysical Journal **676**, 944 (2008).
[47] S. van Velzen, G. R. Farrar, S. Gezari, N. Morrell, D. Zaritsky, L. Ostman, M. Smith, J. Gelfand, and A. J. Drake, Astrophys. J. **741**, 73 (2011), arXiv:1009.1627 [astro-ph.CO].
[48] J. L. Synge, Mon. Not. R. Astron. Soc. **131**, 463 (1966).
[49] J. P. Luminet, Astron. Astrophys. **75**, 228 (1979).
[50] J. Bardeen, Rapidly rotating stars, disks, and black holes, black holes, ed. c. dewitt, bs dewitt (1973).
[51] R. Kumar, A. Kumar, and S. G. Ghosh, Astrophys. J. **896**, 89 (2020), arXiv:2006.09869 [gr-qc].
[52] M. Afrin, S. Vagnozzi, and S. G. Ghosh, Astrophys. J. **944**, 149 (2023), arXiv:2209.12584 [gr-qc].
[53] B. Puliçe, R. C. Pantig, A. Övgün, and D. A. Demir, Classical and Quantum Gravity (2023).
[54] Z. Yan, X. Zhang, M. Wan, and C. Wu, Eur. Phys. J. Plus **138**, 377 (2023), arXiv:2304.07952 [gr-qc].
[55] E. Ghorani, B. Puliçe, F. Atamurotov, J. Rayimbaev, A. Abdujabbarov, and D. Demir, Eur. Phys. J. C **83**, 318 (2023).
[56] F. Atamurotov, I. Hussain, G. Mustafa, and A. Övgün, Chin. Phys. C **47**, 025102 (2023).
[57] V. Perlick and O. Y. Tsupko, Phys. Rep. **947**, 1 (2022), arXiv:2105.07101 [gr-qc].
[58] G. Mustafa, F. Atamurotov, I. Hussain, S. Shaymatov, and A. Övgün, Chin. Phys. C **46**, 125107 (2022).
[59] R. C. Pantig, L. Mastrototaro, G. Lambiase, and A. Övgün, Eur. Phys. J. C **82**, 1155 (2022), arXiv:2208.06664 [gr-qc].
[60] I. Banerjee, S. Sau, and S. SenGupta, JCAP **2022** (09), 066.
[61] N. U. Molla and U. Debnath, Annals Phys. **453**, 169304 (2023), arXiv:2212.02104 [gr-qc].
[62] F. Atamurotov, I. Hussain, G. Mustafa, and K. Jusufi, Eur. Phys. J. C **82**, 831 (2022).
[63] F. Atamurotov, U. Papnoi, and K. Jusufi, Classical and Quantum Gravity **39**, 025014 (2022), arXiv:2104.14898 [gr-qc].
[64] S. V. Ketov, Universe **8**, 351 (2022).
[65] E. Witten, Nuclear Physics B **500**, 3 (1997).
[66] S. Ketov and S. Mochalov, Physics Letters,(Section) B;(Netherlands) **261** (1991).
[67] Z. Elgood, T. Ortín, and D. Pereñíguez, Journal of High Energy Physics **2021**, 1 (2021).
[68] V. R. Ivanov, S. V. Ketov, E. O. Pozdeeva, and S. Y. Vernov, Journal of Cosmology and Astroparticle Physics **2022** (03), 058.
[69] E. Falcón-Gómez *et al.*, Eur. Phys. J. C **82**, 1175 (2022), [Erratum: Eur.Phys.J.C 83, 234 (2023)], arXiv:2303.06085 [gr-qc].
[70] E. Mitsou and J. Yoo, *Tetrad Formalism for Exact Cosmological Observables*, SpringerBriefs in Physics (Springer, 2020) arXiv:1908.10757 [astro-ph.CO].
[71] R. M. Wald, *General relativity* (University of Chicago press, 2010).
[72] L. C. Crispino, A. Higuchi, L. A. Oliveira, and E. S. de Oliveira, The European Physical Journal C **76**, 1 (2016).
[73] U. Papnoi and F. Atamurotov, Physics of the Dark Universe **35**, 100916 (2022), arXiv:2111.15523 [gr-qc].
[74] K. Akiyama and et al., Astrophys. J. Lett **930**, L12 (2022).
[75] C. Bambi, K. Freese, S. Vagnozzi, and L. Visinelli, Phys. Rev. D **100**, 044057 (2019), arXiv:1904.12983 [gr-qc].

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Facial Image Pre-processing for Enhanced Ethnicity Classification

Keshav Goel
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
keshavgoel_2k19co186@dtu.ac.in

Hitesh Dhiman
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
hiteshdhiman_2k19co162@dtu.ac.in

Himanshu Yadav
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
himanshuyadav_2k19co161@dtu.ac.in

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

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205642

***Abstract*—Facial recognition and ethnicity detection is an important task in various applications such as security, marketing, and social sciences. However, the accuracy of Face recognition and ethnicity detection algorithms can be impacted by various factors such as lighting conditions, noise and irrelevant features in image background, variation in facial expressions, face size, pose and orientation across images. In this study, we present a novel face image preprocessing model in this work that aims to enhance the performance of deep learning models like ResNet, DenseNet and other state of the art Deep Learning models on facial recognition tasks, specifically ethnicity detection by addressing these issues. Our novel method for preprocessing facial images incorporates several techniques to enhance the quality of the input images. Firstly, we removed the background using DeepLab-v3. Secondly, we employed facial landmarks detection and facial pose and orientation correction using facial alignment based on the angle between the eyes. Thirdly, to facilitate race classification, we performed facial cropping to isolate the facial region and other critical features essential for the task. Finally, we ensured consistent lighting conditions across all images using the Multi-scale retinex algorithm. These techniques together enabled the deep learning model to better analyze and classify facial data. To assess the efficacy of our model, we carried out a series of experiments on the UTK Face dataset and compared the outcomes with those obtained when our preprocessing model was not utilized.**

The experimental outcomes show that our suggested preprocessing model improves the f1-score of ethnicity detection classifiers by up to 6%. The proposed facial image preprocessing model is efficient, fast, and easy to implement, enabling real-time applications, making it appropriate. Our research highlights the importance of preprocessing techniques in enhancing ethnicity detection performance and provides a practical solution for improving the accuracy of ethnicity detection and facial recognition systems.

***Index Terms*—Ethnicity detection, Facial recognition, Face preprocessing, UTK Face dataset, Face detection, Multi scale retinex illumination, ResNet, EfficientNet, DenseNet, VGG**

I. Introduction

Facial recognition and ethnicity detection have become increasingly important in various applications, including security, marketing, and social sciences. In a recent study [6], it was observed that GANs generate data that belong to the same distribution as the input dataset. Consequently, using GANs [2] for data augmentation would only serve to increase the existing biases in the data. Hence, it is essential to acknowledge the potential for bias in GAN-generated data and to implement strategies to mitigate them. Facial ethnicity detection tasks can be quite challenging owing to the wide range of variations in lighting, positions, orientations, and facial expressions between images. Additionally, the presence of background noise in the images can further complicate the task. These factors can significantly impact the accuracy of ethnicity detection algorithms, leading to errors and misclassifications.

To address these challenges, facial image preprocessing has emerged as a critical step in enhancing the performance of ethnicity detection and facial recognition systems. Facial image preprocessing involves applying a series of techniques to normalize and enhance facial images, making them more suitable for comparison across different subjects and under different conditions.

This paper introduces a novel facial image preprocessing model that aims to improve the accuracy of face detection algorithms, with a specific focus on ethnicity classification models. The proposed model is designed to effectively preprocess facial images and extract key features that are essential for accurate ethnicity classification. We have incorporated multiple preprocessing techniques in our proposed model, such as background removal using DeepLab v3 based on the ResNet 101 model, facial landmarks detection, facial pose and orientation correction using facial alignment based on the angle between the eyes, face cropping, and ensuring consistent lighting conditions across different images using the multi-scale retinex algorithm.

To assess the effectiveness of our proposed facial image preprocessing model, we conducted several experiments on the

979-8-3503-3860-7/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:28 UTC from IEEE Xplore. Restrictions apply.

UTK Face dataset [13] using various different models, comparing the results obtained with and without the application of our preprocessing model. Our experimental findings indicate that the proposed model led to a significant improvement in f1-score of ethnicity detection models, with performance gains of upto 6%. The proposed facial image preprocessing model is efficient, fast, and easy to implement, making it suitable for real-time applications.

This research highlights the importance of preprocessing techniques in enhancing the ethnicity detection performance and provides a practical solution for improving the accuracy of ethnicity detection and facial recognition systems.The rest of the work comprises the following sections: An overview of relevant work in face recognition and facial image preprocessing is provided in Section II. Section III provides a description of the suggested face image preprocessing model and the preprocessing techniques used. Section IV provides the methodology used to evaluate the proposed model. The experimental findings are presented in Section V, along with a comparison of the outcomes obtained with and without the use of our preprocessing. The work is finally concluded in Section VI, which also covers future research.

II. Related Work

In the field of facial recognition and its related tasks, several researchers have investigated the use of preprocessing techniques to enhance the accuracy of these tasks. Deep learning models have achieved remarkable success in feature extraction from the facial images. However, the quality of facial datasets is generally not so good and so it make it difficult for even those modern technology to extract facial traits from the images, using deep learning models. Hence it is important to develop effective and efficient methods and tools to preprocess the images.

In a previous study on illumination preprocessing for facial images, Han et al. [3] proposed a method that considers lighting differences between different faces. The research yielded valuable insights and findings on how to develop more effective illumination preprocessing techniques. They discussed and compared existing methods in the field. A research paper [14] provides a comprehensive survey of face recognition techniques, including preprocessing techniques used to improve face recognition accuracy. The authors discuss various preprocessing techniques, such as illumination normalization, face alignment, and feature extraction, and how they might be applied to improve how well face recognition systems function. A study [11] proposes a deep convolutional neural network (CNN) architecture for ethnicity and gender classification. The authors used a large dataset of facial images and applied various preprocessing techniques, such as cropping, histogram equalization, and scaling, to improve the performance of the CNN. Recent Survey [9] of face detection techniques, including preprocessing techniques used to improve face detection accuracy. The authors discuss various preprocessing techniques, such as illumination normalization, face alignment, and pose estimation, and how they might be applied to improve how well face detection algorithm works.

Facial image preprocessing techniques

In order to increase the precision and effectiveness of face recognition and ethnicity identification algorithms, facial image preprocessing methods are crucial. These are several methods that are often used:

- **Facial background removal:** This technique removes the background of the image to focus only on the facial region. It helps to reduce the influence of the image background on the facial recognition process. [1] In a paper, a novel method of semantic image segmentation is proposed, which entails segmenting an image into many parts and giving each part a label to denote the objects in the picture. The suggested method makes use of atrous convolution, a kind of convolution that permits the inclusion of a broader geographical context without requiring an increase in the number of parameters. The atrous spatial pyramid pooling (ASPP) module, a modified atrous convolutional layer, is a component of the DeepLab v3 architecture that is used to extract information from various scales.
- **Facial landmarks detection:** Facial landmarks detection is used to identify the important facial landmarks in the picture, such as the nose, mouth, and eyes. It helps to align the face to a consistent reference frame, which is crucial for accurate facial recognition.
- **Face alignment:** It adjusts the orientation and pose of the face in the image to a consistent reference frame. It helps to reduce the variation in facial orientation and pose, which can impact the accuracy of facial recognition.
- **Face Cropping:** It involves cropping the facial region of an image, isolating only the face, and removing any extraneous background. This technique can improve facial recognition and ethnicity detection accuracy by reducing the size of the image and removing non-facial features that may interfere with feature detection.
- **Multi scale retinex illumination:** It is a technique for correcting illumination variations across facial images. This technique adjusts the brightness and contrast of facial images to achieve consistent lighting conditions across various images, improving facial recognition and ethnicity detection accuracy by reducing variations in image intensity caused by lighting differences and Camera conditions.

Overall, these preprocessing techniques are necessary to enhance the accuracy and efficiency of facial recognition. It is an important and active area of research, with ongoing efforts to develop effective and efficient methods and tools to preprocess the images.

III. Proposed Model

In the proposed model, We have used the following facial image preprocessing techniques in a specific order to the input images. 1) Facial Background Removal to facilitate face and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:28 UTC from IEEE Xplore. Restrictions apply.

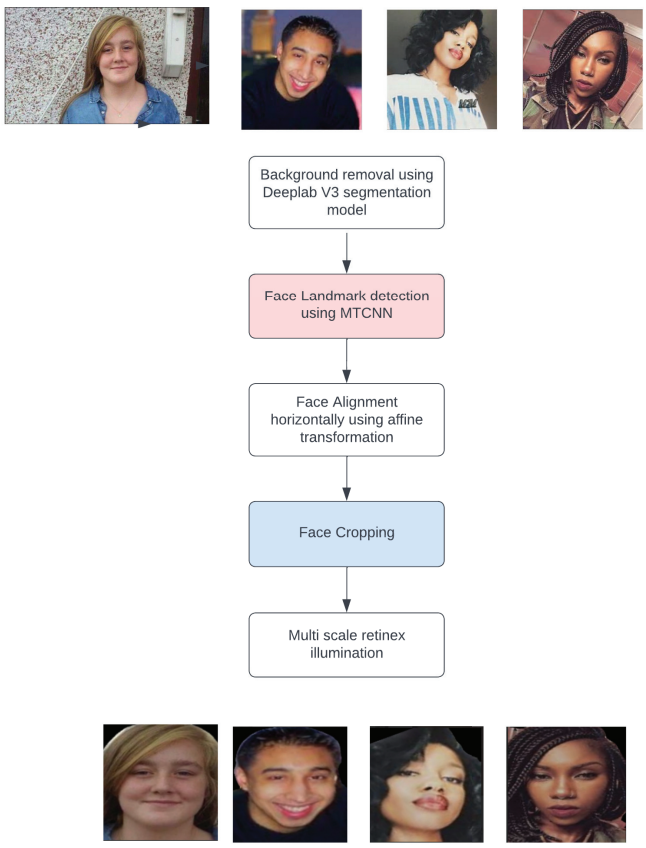


Fig. 1. Flowchart illustrating our image preprocessing approach

facial landmarks detection. 2) Face and Facial Landmarks Detection is applied to detect face and important facial landmarks such as mouth, nose, and eyes. 3) Face Alignment was performed to ensure that all faces are horizontally oriented. This was achieved by first using the facial landmarks detected in the previous step to determine the midpoint between the eyes as a reference point for rotation transformation. Then We computed the angle between the line joining the two eyes and the horizontal axis, which was utilized to ascertain the degree of rotation necessary for alignment. Subsequently, a transformation matrix was created to represent the required rotation and scaling using the midpoint of the two eyes and angle between the horizontal axis and line joining the two eyes. Finally, the transformation matrix was employed to the input image to align the face. 4) Face Cropping, was then employed to crop the face images, including the facial region, hair, and other essential features crucial for ethnicity determination. For this task, the face detection algorithm was applied again at this step to determine the facial region in the aligned image. Subsequently, the region was expanded to enhance the face's capture and facilitate the collection of additional details vital for the task at hand. 5) Multiscale Retinex Illumination [7] was implemented, utilizing the multi-scale Retinex algorithm to adjust the image's illumination to achieve consistent lighting conditions across different images. The multi-scale Retinex algorithm was selected due to its ability to modify illumination while preserving image details.

Implementation

For background removal, we have used the DeepLab v3 model based on ResNet 101. For facial landmarks detection, We have used the MTCNN (Multi-task Cascaded Convolutional Networks) [12] library. For face cropping, the facial region captured using MTCNN was expanded by adding 15 percent of the original width to the left and right sides and 20 percent of the original length to the top and bottom sides. Finally we used Multiscale Retinex algorithm to adjust the illumination of the image to achieve consistent lighting conditions across different images. The target illumination was set to 0.7 for all the images.

IV. Methodology

In this paper for Data Collection and Preparation, we used the UTK Face dataset, which includes annotations for age, gender, and ethnicity, and comprises of more than 20,000 images. The dataset comprises images that exhibit a wide range of characteristics, such as different poses, facial expression, image background, illumination, and resolution. We randomly selected 10,000 images from the dataset and divided them into training and testing sets each containing 5,000 images.

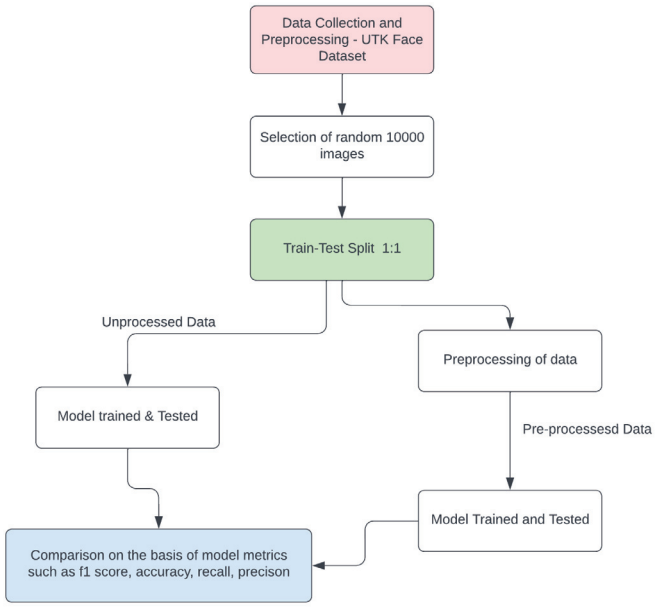


Fig. 2. Flowchart depicting the method used to compare the Deep Learning model with or without images preprocessing with our tool

For preprocessing the training and testing images, we have used our preprocessing model as explained in Section III.

For model training and testing, we used transfer learning [15] to train various state of the art Deep Learning models like ResNet [4], EfficientNet [10], DenseNet [5] and VGG

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:28 UTC from IEEE Xplore. Restrictions apply.

[8] on the unpreprocessed and preprocessed images. To do this, we changed the last layers to train the model using the cross entropy loss function. To train the model for ethnicity classification task, we used the ethnicity annotations given in the dataset.

Then we evaluated the performance of the models with and without using the proposed preprocessing model on the testing set using the following metrics: Recall, F1 Scores, and Precison. We compared the both results for each of the models to evaluate the effectiveness of our proposed preprocessing model.

Implementation

For training we used a cross entropy loss function. Using a 16 as a batch size, and 0.01 as a starting learning rate, the model was trained. To enhance the model's convergence, After each epoch, the learning rate dropped by a factor of 0.1. During the testing phase, we calculated the F1 score, precision, and recall of all the classifiers in order to assess their performance.

V. Results and Discussion

To assess the efficacy of our preprocessing model, we used the metrics like Recall, Precision and overall F1 score, which are widely-used evaluation measures for classification tasks. While accuracy is a commonly used evaluation metric in machine learning, it may not be reliable for imbalanced datasets where one class significantly outweighs the others. In such scenarios, a model that always predicts the majority class may achieve a high accuracy score but may not be practically useful. To address this issue, the F1 score considers both precision and recall, which are metrics used to evaluate a model's classification accuracy. Recall is the percentage of properly recognised positive occurrences among all actually occurring positive occurrences, precision is the percentage of correctly identified positive occurrences among all forecasted positive occurrences. The F1 score, which ranges from 0(worst) to 1(best), is the harmonic mean of recall and accuracy. F1 score (1) is the equation to determine the F1 score. Given that the F1 score indicates the balance between recall and precision, it is particularly helpful for datasets with imbalances while accuracy gauges overall performance. Therefore, we used Recall, Precision, and F1 score to provide a more comprehensive evaluation of our preprocessing model's effectiveness.

$$\mathcal{F}1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{1}$$

ResNet, DenseNet, EfficientNet, and VGG, four well-known deep learning models, were examined in this study for their performance on facial data with and without our preprocessing tool. The results of the models without using preprocessing model are presented in Table 1, while the results of the models with using preprocessing are presented in Table 2. The confusion matrices for each model without and with preprocessing are shown in Figures 3-6. Our results indicate that using our preprocessing tool significantly improved the performance of deep learning models on facial data. Without preprocessing, ResNet, for instance, achieved a recall, precision, and F1 scores of 0.75, 0.72 and 0.72, respectively. After preprocessing, these scores increased to 0.81, 0.79, and 0.80. The Precision, F1 Score, and Recall for the DenseNet model also increased from 0.80, 0.79, and 0.79, respectively, without preprocessing to 0.83, 0.82, and 0.82 after preprocessing. Without preprocessing, EfficientNet's F1 Score, recall, and precision accuracy were 0.79, 0.79, and 0.80 respectively, but they increased to 0.81, 0.80, and 0.81 after preprocessing. When preprocessing was used, the precision, F1 score and recall accuracy for the VGG model increased from 0.76, 0.75, and 0.74 to 0.78, 0.78, and 0.77, respectively.The outcomes in Tables 1 and 2 demonstrate the performance of deep learning models on facial data may be greatly enhanced by the use of our preprocessing method. The models which are having a low f1 scores have gone more pronounced increases compared to models with greater scores. VGG model have lowest score which is due to the fact that is is unable to transfer information of upper layer to lower layers. The results also indicate that the use of our preprocessing model for facial image data can enhancing deep learning models capacity to capture important features in the image. Specifically, we observed a significant improvement in model performance when the images were processed using our tool compared to when they were not. This suggests that our preprocessing model can facilitate better feature extraction and enhance the general performance of facial data-based deep learning models.

Fig. 3. ResNet Model Confusion Matrix (a) without preprocessing (b) with preprocessing

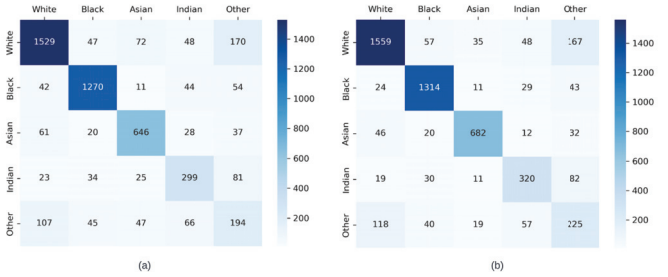


Fig. 4. DensNet Model Confusion Matrix(a) without preprocessing (b) with preprocessing

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:28 UTC from IEEE Xplore. Restrictions apply.

TABLE I
MODELS WITH RESULTS: WITHOUT PREPROCESSING

Models	Precision	Recall	f1 Score
ResNet	0.7232	0.7494	0.7249
DenseNet	0.7967	0.7876	0.7916
EfficientNet	0.7959	0.7886	0.7914
VGG	0.7588	0.7426	0.7462

TABLE II
MODELS WITH RESULTS: WITH PREPROCESSING

Models	Precision	Recall	f1 Score
ResNet	0.7930	0.812	0.7979
DenseNet	0.8287	0.82	0.8236
EfficientNet	0.8123	0.8024	0.8066
VGG	0.7836	0.7708	0.7759

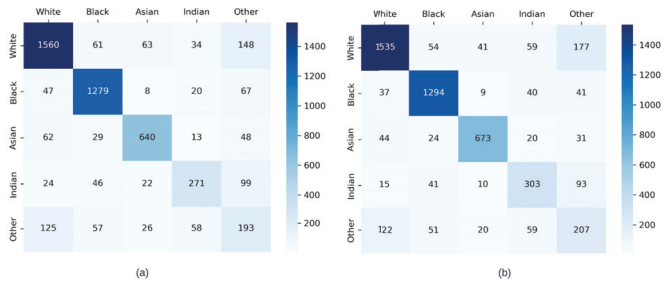


Fig. 5. EfficientNet Model Confusion Matrix (a) without preprocessing (b) with preprocessing

Fig. 6. VGG Model Confusion Matrix (a) without preprocessing (b) with preprocessing

We also have observed near consistent trends across different models in terms of class performance. Across all models, black and white races outperformed the other races due to their more distinct features, which make them more easily identifiable. Asian individuals also performed well, albeit slightly worse than black and white individuals. However, Indians performed relatively poorly, coming in fourth place, due to the significant variability in their skin tones, facial features, and face sizes. They often share more features with other races, making them less distinguishable from others. As for class Other, it generally exhibited poor performance across all the models, as it consist of two or three different races, which may be less similar to each other and more similar to races in remaining classes. For instance, Middle Eastern people which are in class others closely resemble white people, while Latino and Hispanic people which are in class others closely resemble Indians and Blacks.

VI. CONCLUSION AND FUTURE WORK

We developed a unique preprocessing model in this work that is designed to cater to the unique characteristics of facial images. Our research demonstrates the significant impact of facial image preprocessing on the performance of deep learning models in ethnicity classification tasks. The results of our experiments show that the use of our preprocessing model significantly improved the performance of popular deep learning models such as ResNet, DenseNet, EfficientNet, and VGG where processed images were utilized for training and testing compared to the case when trained and tested on unprocessed images. This highlights the importance of appropriate preprocessing techniques for achieving optimal performance in deep learning tasks related to facial recognition and ethnicity classification. In general, this study makes a significant addition to the field of face recognition by introducing a novel method for preprocessing raw images before feeding to Deep Learning models that can be applied in a variety of face recognition related tasks like ethnicity classification, age estimation and gender classification. The proposed preprocessing model can be a useful tool for researchers and practitioners working on various kind of face recognition related tasks, enabling them to find high-quality preprocessed facial images more effectively.

Our study has demonstrated the effectiveness of the facial image preprocessing methods in enhancing the quality of facial images and improving the performance of deep learning models in ethnicity detection tasks. However, there are still several areas where further research could extend our findings and improve the utility of our methods. Our study showed the improve in accuracy of deep learning models in ethnicity classification task. Future research could explore the applicability of our proposed preprocessing tool for other

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:28 UTC from IEEE Xplore. Restrictions apply.

face recognition tasks such as age estimation and gender classification. Our study employed a relatively small dataset for evaluation purposes. Future research could evaluate the effectiveness of our methods on larger and more diverse datasets. This could provide more robust evidence of the generalizability of our methods and their performance in a wider range of applications. Future researchers could also try to investigate additional preprocessing techniques that can be integrated into the model to further improve its performance.

Furthermore, our study also sheds light on the challenges and limitations of ethnicity classification using facial images. Our analysis indicates that the performance of models varies depending on the ethnic group being classified, with some groups being more difficult to distinguish than others due to the similarity of their facial features. These observations highlight the need for further research in developing more robust and accurate models for ethnicity classification.

REFERENCES

[1] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.

[2] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.

[3] Hu Han, Shiguang Shan, Xilin Chen, and Wen Gao. A comparative study on illumination preprocessing in face recognition. *Pattern Recognition*, 46(6):1691–1699, 2013.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[6] Vongani H Maluleke, Neerja Thakkar, Tim Brooks, Ethan Weber, Trevor Darrell, Alexei A Efros, Angjoo Kanazawa, and Devin Guillory. Studying bias in gans through the lens of race. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII*, pages 344–360. Springer, 2022.

[7] Ana Belén Petro, Catalina Sbert, and Jean-Michel Morel. Multiscale Retinex. *Image Processing On Line*, pages 71–88, 2014. https://doi.org/10.5201/ipol.2014.107.

[8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[9] Ankit Srivastava, Suraj Mane, Aaditya Shah, Nirmit Shrivastava, and Bhushan Thakare. A survey of face detection algorithms. In *2017 International Conference on Inventive Systems and Control (ICISC)*, pages 1–4, 2017.

[10] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.

[11] Wei Wang, Feixiang He, and Qijun Zhao. Facial ethnicity classification with deep convolutional neural networks. In Zhisheng You, Jie Zhou, Yunhong Wang, Zhenan Sun, Shiguang Shan, Weishi Zheng, Jianjiang Feng, and Qijun Zhao, editors, *Biometric Recognition*, pages 176–185, Cham, 2016. Springer International Publishing.

[12] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.

[13] Song Yang Zhang, Zhifei and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[14] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, dec 2003.

[15] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:28 UTC from IEEE Xplore. Restrictions apply.

KSME

Springer

Journal of Mechanical Science and Technology 37 (8) 2023

Original Article
DOI 10.1007/s12206-023-0717-x

Keywords:
· Steam turbine blade
· Scanning electron microscopy (SEM)
· Corrosion pits
· Fatigue striations

Correspondence to:
Nitin Sehra
nitinsehra441@gmail.com

Citation:
Sehra, N., Rani, S. (2023). Failure investigations of last stage low-pressure steam turbine blade. Journal of Mechanical Science and Technology 37 (8) (2023) ?~?.
http://doi.org/10.1007/s12206-023-0717-x

Received November 3rd, 2022
Revised April 23rd, 2023
Accepted May 8th, 2023

† Recommended by Editor Chongdu Cho

© The Korean Society of Mechanical Engineers and Springer-Verlag GmbH Germany, part of Springer Nature 2023

Failure investigations of last stage low-pressure steam turbine blade

Nitin Sehra and Sushila Rani

Department of Mechanical, Production, Industrial and Automobile Engineering, Delhi Technological University, Delhi 110042, India

Abstract This research study examines the failure of the last stage low-pressure steam turbine blade. The blade is made up of martensitic steel of X10CR alloy. The main focus of this paper is to analyze the fracture that appears on the leading edge of the blade at a distance of 350 mm from the blade tip. The investigation includes a visual examination, SEM fractography, microstructure characterization & mechanical testing. On Investigation, silicon oxide (Sio_2) is detected on the cracked part of the blade along with sodium, potassium, chlorides, oxide, silicate & sulfides inside the corrosion pits on the fracture surface. Further, an Intergranular fracture followed by fatigue striations is observed on the blade's leading edge. It is concluded that the failure of the low-pressure steam turbine blade may have been caused by corrosion fatigue.

1. Introduction

Low-pressure steam turbine blades are essential in power generation systems that convert steam energy into electrical power. These blades are designed to withstand extreme heat and pressure, which makes them critical to conduct the smooth operation of a power plant. In a steam turbine, there are multiple stages through which steam passes before it enters the turbine, including HP (high pressure), IP (intermediate pressure), and LP (low pressure) zones [1]. Steam turbine blades are more prone to failure at LP than the HP and IP blades, as LP blades capture the major part of the energy from the flowing steam compared to IP and HP blades [2]. However, despite their robust construction, turbine blades are susceptible to various forms of damage, which can lead to failure and, in some cases, catastrophic consequences from repeated stress cycles due to operational or loading conditions, leading to blade cracking or fracture [3-5].

Low-pressure turbine blades may fail due to various factors, including mechanical stress [6], corrosion [7-9], erosion [10-12], fatigue [13-15], and vibration [16-19]. Mechanical stress may arise due to excessive vibration, misalignment, or improper balancing, leading to blade deformation, creep, thermal distortion, which results in blade cracking. Corrosion may occur due to the presence of contaminants in the steam or condensate, resulting in blade pitting or corrosion fatigue. Erosion can occur due to the impact of solid particles or droplets in the steam, leading to blade surface erosion or wear, which results in fatigue failure of the blade.

Recent research has focused on understanding the mechanisms of failure in steam turbine blades mainly made up of martensitic steel, particularly emphasizing the role of corrosion pits, fatigue cracks, intergranular failure, and corrosion fatigue failure. Corrosion fatigue is a significant factor contributing to the failure of low-pressure steam turbine blades in industrial applications. Corrosion fatigue is a form of material degradation caused by the combined action of mechanical stress and corrosion. Corrosion pits developing on the material's surface indicate the initiation of the corrosion fatigue process, which acts as stress concentrators. As the material is subjected to cyclic loading, cracks develop at the site of the corrosion pits, which can eventually propagate through the material, leading to its failure.

Kim [20] investigated the root causes of crack formation in the fourth-stage blade of a low-pressure steam turbine. The study utilized non-destructive testing techniques, such as ultra-

sonic testing and visual inspection, to evaluate the cracks' extent and severity. The findings suggest that a combination of high-temperature exposure and mechanical stress were the main reasons for crack formation. Also, they have recommended implementing preventive measures, such as improving the blade design and manufacturing process in order to reduce the risk of crack formation. Bhagi et al. [21] analysed the failed L-1 low-pressure steam turbine blade to understand its failure mechanism. They observed that cyclic stresses and environmental effects accelerate corrosion fatigue. They have concluded that corrosion fatigue is the prime reason for blade failure, as evidenced by the transgranular cleavage fracture. Katinic and Kozak [22] presented the fracture investigations of the two adjacent rotor blades, which failed at their roots & observed that the blades failed due to corrosion fatigue. They redesigned the blade to increase their fatigue life. Wei et al. [23] presented the corrosion fatigue cracking mechanism of the 17-4 PH blade of a low-pressure rotor of a steam turbine. They discovered that in a steam environment, potassium ions (K+) and chlorine ions (Cl-) interact physically, chemically, and electrochemically with the turbine blades to produce localized spot corrosion and corrosion pits, which act as stress concentrators & found that the blade cracks formed due to corrosion fatigue. Khan and Sasikumar [24] investigated the failed low-pressure steam turbine blade made of AISI 420 steel. They observed shallow corrosion pits near the crack initiation point and varying loads increased the vibration in the turbine, which caused the fatigue cracks in the weaker zones, leading to the catastrophic failure of the turbine blade. Rivaz et al. [25] examined the 410 stainless steel failed blade after 165000 hours of working. Findings revealed that the inclusion of foreign particles like silicon oxide and ferric oxide in steam could cause corrosion in blades, and corrosion fatigue is the main reason for failure. Rani and Agrawal [26] presented the failure investigations of 210 MW last-stage low-pressure steam turbine blades after 152241 hours of working. They have analysed the blade through mechanical analysis, metallurgical analysis, and Non-destructive testing (NDT) and found that near the edges of the blade, water-droplet erosion creates erosion pits, which leads to the failure of the blade.

After a comprehensive literature review, it is observed that preventive measures have to be taken to reduce the crack formation and risk of blade failure resulting in enhancing the fatigue life of the blade. There are different methods to enhance the blade's fatigue life, like shot peening [27-29], coating the blade [30, 31], and providing dampers between the blades [32] in order to reduce the failure rate of the blade.

In this research work, failure analysis of a low-pressure steam turbine blade of a 210 MW thermal power plant is presented. High levels of noise and vibration at the LP zone during operation caused the unit to shut down and one of the blades from the 29th stage was found to be fractured on opening the turbine casing. The fractured blade is made of X10CrNiMoV12-2-2 alloy. The investigations include visual observations, mechanical tests, and scanning electron microscopy (SEM) accompanied by EDS (energy dispersive spectroscopy).

2. Experimental details

The last stage steam turbine blade of the L-0 row is collected and investigated fractographically to determine the main reasons for its failure. A transverse crack was noticed on the blade's airfoil-leading edge at a distance of 350 mm from the blade's tip, as shown in Fig. 1. The samples were prepared from the cracked region of the blade as shown in Fig. 2. The specimen was prepared using standard metallographic techniques for fractographic analysis, and the prepared sample's microstructure was revealed using Glycergia etchant through an optical microscope. For crack morphology, scanning electron microscopy of the prepared sample is carried out on VEGA3,

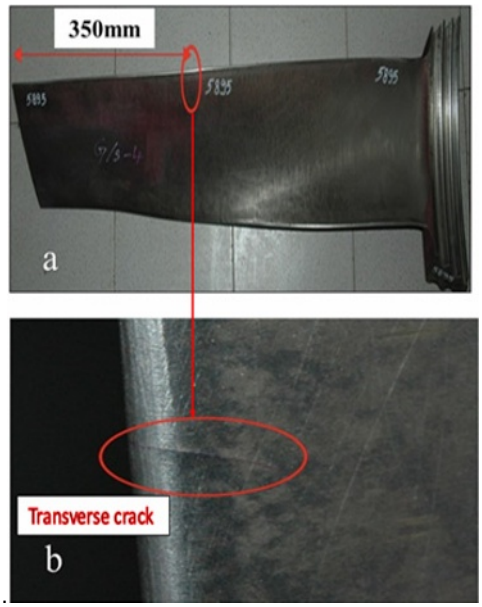


Fig. 1. Transverse crack on the airfoil-leading edge of the steam turbine blade.

Fig. 2. Sample cut at the cracked region for fractographic investigation.

TESCAN model accompanied with energy dispersive spectroscopy (EDS). Further, the hardness of the blade material was evaluated using the Vickers hardness test at a force of 0.5 kg for 10 seconds. And a tensile test was conducted to assess the mechanical properties of the blade material.

3. Results & discussion

3.1 Visual observations

On visual examination, a transverse crack of around 20-25 mm is observed at a distance 350 mm away from the blade's tip, which propagates from the leading edge towards the trailing edge as shown in Fig. 1. Regions of black patches i.e. iron oxide (Fe_3O_4) are also found on the blade's surface from tip to root of the blade as shown in Fig. 3 and presence of iron oxide (Fe_3O_4) can make the blade material weak and ductile. Further, water droplet erosion creates erosion pits on the airfoil-leading edge but these are away from the transverse crack as shown in Fig. 4, which concludes that the erosion pits are not the reason for the failure of the blade.

3.2 Chemical analysis

An optical emission analyzer was used to identify the chemical makeup of the blade material. The result of the chemical composition of X10CRNIMOV 12-2-2 is tabulated in Table 1. Chemical testing results reveal that the turbine blade material corresponds with AISI 410-grade martensitic steel.

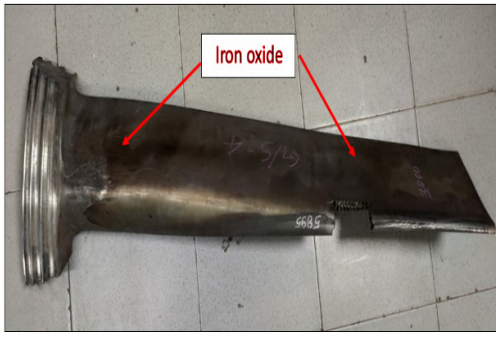


Fig. 3. Black patches iron oxide (Fe_3O_4) observed on the blade's surface.

Fig. 4. Erosion pits at the blade airfoil-leading edge.

3.3 Microstructure analysis

Optical microscope is used to identify the microstructure of the blade material. Figs. 5(a) and (b) presented the microstructures of blade material at 100x and 200x respectively. In the observed microstructures hardened and tempered martensite is uniformly distributed and showed no signs of microstructural deterioration. The blade specification corresponds with the X10CRNIMOV12-2 alloy.

3.4 SEM analysis

For scanning electron microscopy (SEM) the specimen was prepared from the cracked region of the steam turbine blade as shown in Fig. 6. SEM of the specimen was carried out by using a VEGA3 TESCAN MODEL equipped with an EDS facility. For fractographic analysis, the specimen is divided into three regions viz. R1, R2 & R3 as shown in Fig. 6(b). The fractographic results of region R1 are shown in Fig. 7, which depicts corrosion pits. The environment-assisted cracking of LP steam turbine blades usually begins from corrosion pits. The corrosion pits act as stress concentrators or stress raisers; When a blade is in operation, it experiences cyclic loading due to the rotational forces and the steam flow. These cyclic loads can cause small cracks to initiate at the corrosion pits due to the local stress concentrations. Over time, these cracks can grow and eventually lead to a catastrophic failure of the blade. The presence of corrosion pits can also reduce the fatigue life

Table 1. Composition of turbine blade material.

Composition	C	Cr	Mo	Ni	V	Mn	Si
Measured weight-%	0.21	12.69	1.95	2.73	0.32	0.97	0.46

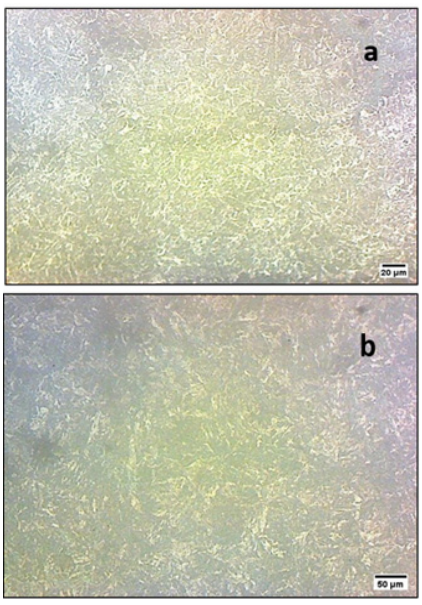


Fig. 5. Optical micrographs showing the tempered martensitic microstructure: (a) microstructure at 100x; (b) microstructure at 200x.

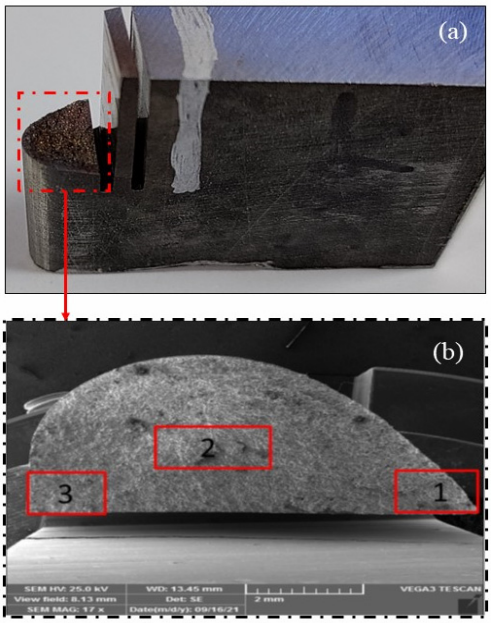


Fig. 6. (a) Macro-morphology of the cracked part of the blade; (b) micro-morphology or SEM view of the rectangular portion divided into three different regions R1, R2, and R3 for fractographic analysis.

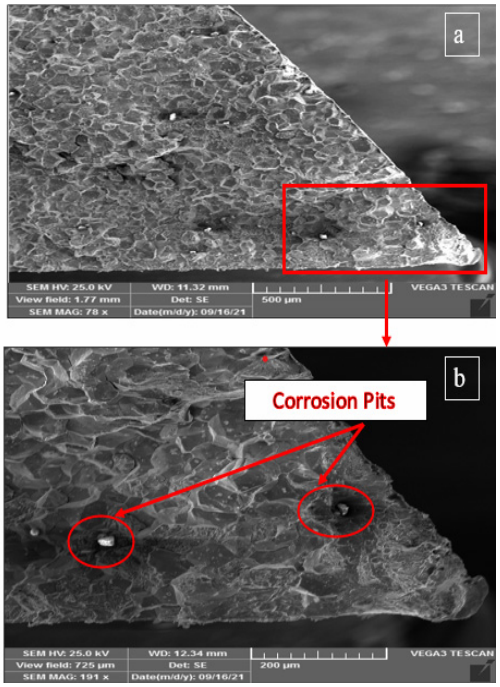


Fig. 7. SEM fractography of R1 region depicts corrosion pits: (a) normal view; (b) the magnified view of the corrosion part.

of the blade by reducing the thickness of the material at the pit. This reduction in thickness can increase the stress level in the material at the pit, making it more susceptible to crack initiation and propagation.

The SEM analysis result of the R2 region shown in Fig. 8, depicts the intergranular failure mode. The results of the SEM

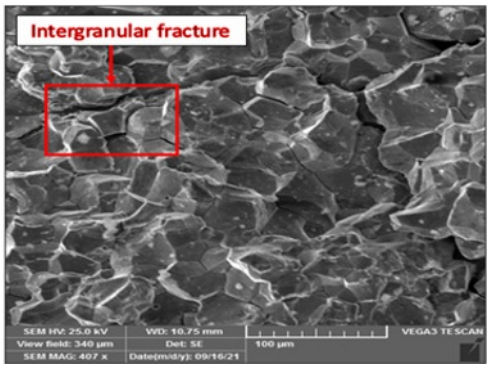


Fig. 8. SEM fractography of region 2 showing the intergranular fracture.

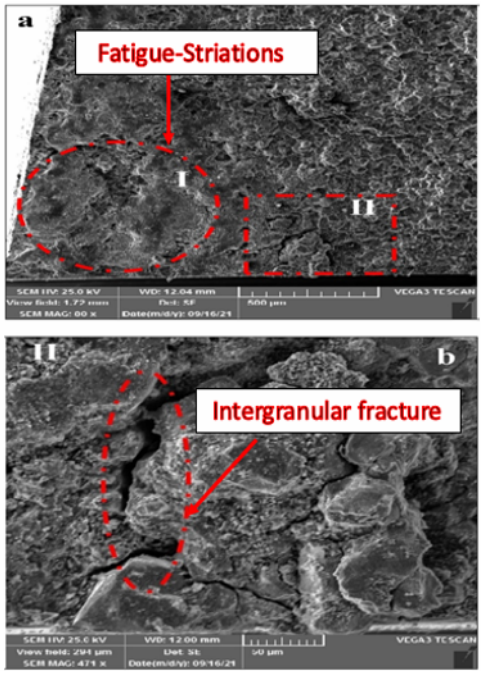


Fig. 9. (a) SEM analysis of the R3 region represents the beach marks and fatigue striations followed by the intergranular fracture; (b) the magnified view of intergranular fracture.

analysis of the R3 region are shown in Figs. 9(a) and (b), representing the beach marks and fatigue striations followed by the intergranular fracture. Fatigue striations are microscopic lines or ridges that form on the surface of a material as a result of repeated cycles of stress. Fatigue striations and intergranular fractures on a low-pressure steam turbine blade can be caused by a number of factors, including material defects, stress concentrations, and cyclic loading. The fatigue striations also show that the crack propagates toward the blade's trailing edge.

3.5 EDS analysis

EDS analysis was performed on the R1 region as shown in Fig. 12. The results of the EDS analysis confirm the presence of chlorine, silicon, oxygen, and other corrosion-promoting products. Silicon and oxygen combine together to form SiO_2

Table 2. Hardness values (HV) at a different location on the blade surface.

Distance from the fracture surface (mm)	Hardness (HV)
0.5	492.43 HV0.5
1	466.63 HV0.5
1.5	416.89 HV0.5
2	395.08 HV0.5

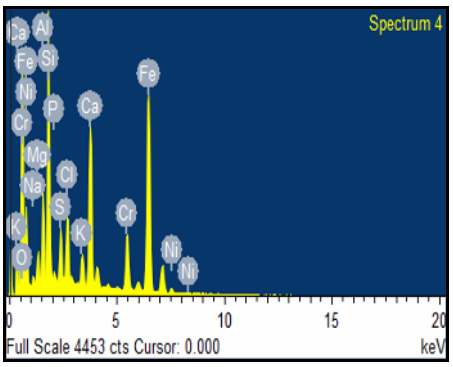


Fig. 10. EDX analysis of the fractured surface of the blade for the R1 region.

particles, which repeatedly strike the blade material and form deposits that create grooves/pits in the blade material and distort the steam passage. The pits penetrated the material, causing localized stress concentrations leading to fatigue cracks. The presence of foreign particles in the steam further exacerbated the problem by providing additional stress concentrations. The corrosion fatigue cracks frequently initiate from corrosion pits, especially when chlorine and oxygen are present & propagate under cycling loading. It is clear from the above findings that corrosion fatigue might be the reason for blade failure.

3.6 Hardness

A hardness test was conducted on a Vickers hardness testing machine near the failed region of the blade. A force of 0.5 kg was applied on the surface of the blade material for 10 seconds as shown in Fig. 11. The results of the hardness test are tabulated in Table 2. And plotted graphically as shown in Fig. 12. It is observed that the hardness is significantly higher on the blade's fracture surface. The increase in hardness of the blade material may be due to strain hardening, Improper heat treatment, contamination & aging factor that can make the material more brittle and could potentially reduce the fatigue strength of the blade material that leads the blade more prone to cracking or fracture under stress or cyclic loading.

3.7 Tensile test

Tensile tests were conducted on the blade's failed and unfailed regions of the blade. The stress-strain curve is shown in Fig. 13. The unfailed region of the blade has a yield strength of 530 MPa and an ultimate tensile strength of 865 MPa, while

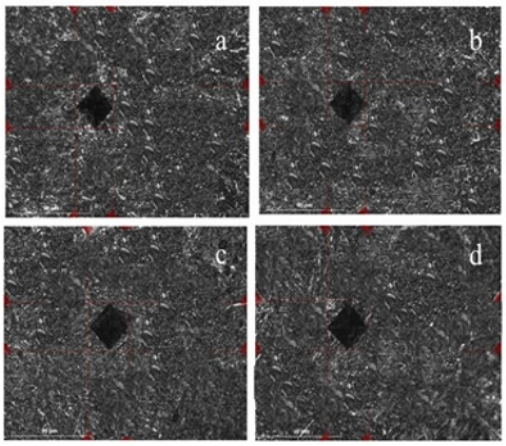


Fig. 11. Vickers hardness test of the blade material at a force of 0.5 kg for 10 seconds.

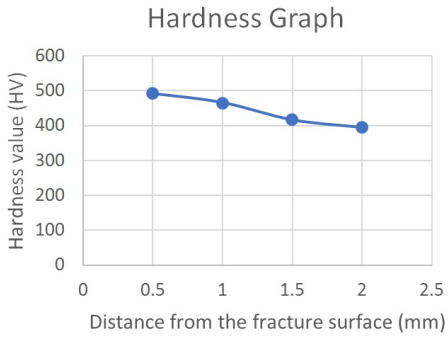


Fig. 12. Graphical plot of hardness value vs distance from the fracture surface.

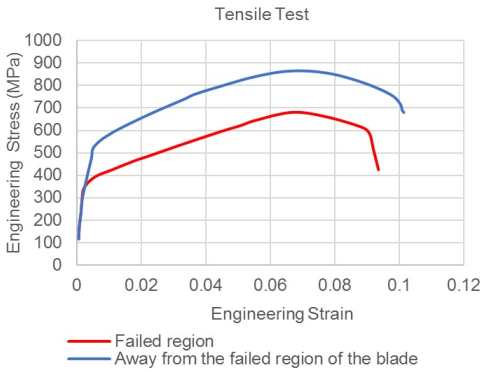


Fig. 13. Stress-strain curve for the failed and the unfailed region of the blade.

the failed region has a yield strength of 343 MPa and an ultimate tensile strength of 680 MPa. The reason for the lower value of the failed region is as it has undergone significant deformation and damage & which reduces the strength and ductility of the material.

4. Conclusions and suggestions

1) The microstructure observed by an optical microscope re-

vealed that hardened and tempered martensite blade microstructure is uniformly distributed and shows no signs of microstructural deterioration. From the findings it is determined that there is no material flaw that causes the turbine blade to fail.

2) Water droplet erosion produces erosion pits on the leading edge of the blade, but they are not near the transverse crack and cannot be considered the cause of the blade's failure.

3) The results of mechanical tests showed that the material's mechanical properties had deteriorated, with the material being more brittle due to an increase in hardness and a fall in ultimate tensile and yield stresses, respectively.

4) The presence of foreign particles and impurities in the steam such as chlorine, silicon, oxygen, and other corrosion-promoting products triggers the failure mechanism.

5) SiO_2 particles repeatedly strike the blade material and form deposits that create grooves/pits in the blade material and distort the steam passage. These pits can act as stress concentrators and increase the risk of fatigue failure.

6) The corrosion pits develop the region of stress-raiser as they form the stress concentration region & when the material is subjected to cyclic loading, cracks develop at the site of the corrosion pits. The crack is initiated & propagated towards the trailing edge. The intergranular fracture occurred, followed by fatigue striations. It has been determined that the failure of the steam turbine blade was attributed to corrosion fatigue failure.

5. Suggestions

1) The chance of pitting is increased by steam contamination which contains dissolved salts, chlorides, or other impurities and could raise the corrosion potential. Regular monitoring of steam quality can help detect contamination early. This can include the measurement of parameters such as conductivity, pH, and dissolved oxygen levels.

2) To enhance the fatigue life of low-pressure steam turbine blades. It is advised to use an appropriate coating or surface modification treatment, such as shot peening or metal-matrix composite coatings. MMCs can be typically applied using a thermal spray process, such as plasma spray or high-velocity oxy-fuel (HVOF) spray.

Nomenclature

LP : Low-pressure
CF : Corrosion fatigue
MPa : Mega pascal

References

[1] K. M. Perkins and M. R. Bache, The influence of inclusions on the fatigue performance of a low pressure turbine blade steel, *Int. J. Fatigue*, 27 (6) (2005) 610-616, doi: 10.1016/j.ijfatigue.2004.12.004.

[2] S. Saxena, J. P. Pandey, R. S. Solanki, G. K. Gupta and O. P. Modi, Coupled mechanical, metallurgical and FEM based failure investigation of steam turbine blade, *Eng. Fail. Anal.*, 52 (2015) 35-44, doi: 10.1016/j.engfailanal.2015.02.012.

[3] G. Das, S. Ghosh Chowdhury, A. Kumar Ray, S. Kumar Das and D. Kumar Bhattacharya, Turbine blade failure in a thermal power plant, *Eng. Fail. Anal.*, 10 (1) (2003) 85-91, doi: 10.1016/S1350-6307(02)00022-5.

[4] Z. Mazur, R. Garcia-Illescas, J. Aguirre-Romano and N. Perez-Rodriguez, Steam turbine blade failure analysis, *Eng. Fail. Anal.*, 15 (1-2) (2008) 129-141, doi: 10.1016/j.engfailanal.2006.11.018.

[5] S. Rani, A. K. Agrawal and V. Rastogi, Failure analysis of a first stage IN738 gas turbine blade tip cracking in a thermal power plant, *Case Stud. Eng. Fail. Anal.*, 8 (2017) 1-10, doi: 10.1016/j.csefa.2016.11.002.

[6] J. Kim, Fatigue failure analysis of last stage blade in a low-pressure steam turbine, *Engineering Failure Analysis*, 6 (2) (1999) 93-100, https://doi.org/10.1016/S1350-6307(98)00034-X.

[7] R. Ebara, Corrosion fatigue crack initiation behavior of stainless steels, *Procedia Eng.*, 2 (1) (2010) 1297-1306, doi: 10.1016/j.proeng.2010.03.141.

[8] K. M. Perkins and M. R. Bache, Corrosion fatigue of a 12 %Cr low pressure turbine blade steel in simulated service environments, *Int. J. Fatigue*, 27 (10-12) (2005) 1499-1508, doi: 10.1016/j.ijfatigue.2005.06.031.

[9] L. Bin Niu, K. Okano, S. Izumi, K. Shiokawa, M. Yamashita and Y. Sakai, Effect of chloride and sulfate ions on crevice corrosion behavior of low-pressure steam turbine materials, *Corros Sci.*, 132 (2018) 284-292, doi: 10.1016/j.corsci.2017.12.017.

[10] C. R. F. Azevedo and A. Sinátora, Erosion-fatigue of steam turbine blades, *Eng. Fail. Anal.*, 16 (7) (2009) 2290-2303, doi: 10.1016/j.engfailanal.2009.03.007.

[11] M. Ahmad, M. Schatz and M. V. Casey, Experimental investigation of droplet size influence on low pressure steam turbine blade erosion, *Wear*, 303 (1-2) (2013) 83-86, doi: 10.1016/j.wear.2013.03.013.

[12] M. S. Khan and C. Sasikumar, A water droplet erosion-induced fatigue crack propagation and failure in X20Cr13 martensitic stainless-steel turbines working at low pressure, *Eng. Fail. Anal.*, 139 (2022) doi: 10.1016/j.engfailanal.2022.106491.

[13] V. N. Shlyannikov, R. R. Yarullin and A. P. Zakharov, Fatigue of steam turbine blades with damage on the leading edge, *Procedia Materials Science*, 3 (2014) 1792-1797, doi: 10.1016/j.mspro.2014.06.289.

[14] S. Cano et al., Detection of damage in steam turbine blades caused by low cycle and strain cycling fatigue, *Eng. Fail. Anal.*, 97 (2019) 579-588, doi: 10.1016/j.engfailanal.2019.01.015.

[15] J. A. Rodríguez et al., Fatigue of steam turbine blades at resonance conditions, *Eng. Fail. Anal.*, 104 (2019) 39-46, doi: 10.1016/j.engfailanal.2019.05.027.

[16] A. Shukla and S. P. Harsha, Vibration response analysis of last stage LP turbine blades for variable size of crack in root, *Procedia Technology*, 23 (2016) 232-239, doi: 10.1016/j.protcy.2016.03.022.

[17] W. Zhao, Y. Li, M. Xue, P. Wang and J. Jiang, Vibration analysis for failure detection in low pressure steam turbine

blades in nuclear power plant, *Eng. Fail. Anal.*, 84 (2018) 11-24, doi: 10.1016/j.engfailanal.2017.10.009.

[18] K. R. Shetkar and J. Srinivas, Localized damage identification in the last stage low-pressure steam turbine blade using dynamic parameter measurements, *J. Nondestruct Eval Diagn Progn. Eng. Syst.*, 6 (1) (2023) doi: 10.1115/1.4056312.

[19] M. Sanvito, E. Pesatori, N. Bachschmid and S. Chatterton, *Analysis of LP Steam Turbine Blade Vibrations: Experimental Results and Numerical Simulations,* Woodhead Publishing Limited (2012) doi: 10.1533/9780857094537.4.189.

[20] H. Kim, Crack evaluation of the fourth stage blade in a low-pressure steam turbine, *Eng. Fail. Anal.*, 18 (3) (2011) 907-913, doi: 10.1016/j.engfailanal.2010.11.004.

[21] L. K. Bhagi, P. Gupta and V. Rastogi, Fractographic investigations of the failure of L-1 low pressure steam turbine blade, *Case Stud. Eng. Fail. Anal.*, 1 (2) (2013) 72-78, doi: 10.1016/j.csefa.2013.04.007.

[22] M. Katinić and D. Kozak, Steam turbine moving blade failure caused by corrosion fatigue - case history, *Procedia Structural Integrity*, 13 (2018) 2040-2047, doi: 10.1016/j.prostr.2018.12.211.

[23] Y. Wei et al., Analysis on corrosion fatigue cracking mechanism of 17-4PH blade of low pressure rotor of steam turbine, *Eng. Fail. Anal.*, 118 (2020) doi: 10.1016/j.engfailanal.2020.104925.

[24] M. Sarim Khan and C. Sasikumar, Failure analysis of AISI 420 steel turbine blade operating at low-pressure, *Mater. Today Proc.*, 66 (2022) 3804-3808, doi: 10.1016/j.matpr.2022.06.197.

[25] A. Rivaz, S. H. Mousavi Anijdan and M. Moazami-Goudarzi, Failure analysis and damage causes of a steam turbine blade of 410 martensitic stainless steel after 165000 h of working, *Eng. Fail. Anal.*, 113 (2020) doi: 10.1016/j.engfailanal.2020.104557.

[26] P. Rani and A. K. Agrawal, Failure analysis of a low-pressure stage steam turbine blade, *Nondestructive Testing and Evaluation*, 38 (4) (2023) 668-682, doi: 10.1080/10589759.2022.2156503.

[27] M. N. James, M. Newby, D. G. Hattingh and A. Steuwer, "Shot-peening of steam turbine blades: residual stresses and their modification by fatigue cycling, *Procedia Engineering*, 2 (1) (2010) 441-451, doi: 10.1016/j.proeng.2010.03.048.

[28] B. K. Pant et al., Studies towards development of laser peening technology for martensitic stainless steel and titanium alloys for steam turbine applications, *Materials Science and Engineering A*, 587 (2013) 352-358, doi: 10.1016/j.msea.2013.08.074.

[29] C. You, B. Y. He, M. Achintha and P. A. S. Reed, Numerical modelling of the fatigue crack shape evolution in a shot-peened steam turbine material, *Int. J. Fatigue*, 104 (2017) 120-135, doi: 10.1016/j.ijfatigue.2017.07.017.

[30] B. C. Navinesh, B. Somasundaram, J. Jagadeeswaran and M. P. Mamatha, Better wear resistance of HVOF coating CrC-NiCrFeSiBCoC (35 %-65 %) and CrC-NiCrFeSiBCoC (80 %-20 %) on SS316 steels, *Materials Today: Proceedings*, 20 (2020) 155-160, doi: 10.1016/j.matpr.2019.10.094.

[31] Z. A. Abdulwahab, S. A. Ajeel and S. I. Jafar, Influence of laser cladding on behavior of fatigue and fatigue corrosion, *IOP Conference Series: Earth and Environmental Science*, 961 (1) (2022) doi: 10.1088/1755-1315/961/1/012035.

[32] L. K. Bhagi, V. Rastogi and P. Gupta, Study of corrosive fatigue and life enhancement of low pressure steam turbine blade using friction dampers, *Journal of Mechanical Science and Technology*, 31 (1) (2017) 17-27, doi: 10.1007/s12206-016-1203-5.

Nitin Sehra is a Ph.D. research scholar in Machine Design in the Mechanical, Production, Industrial and Automobile Engineering Department at Delhi Technological University, Delhi, India. His research area is fatigue failure analysis of low-pressure steam turbine blades.

Sushila Rani has her Ph.D. in Machine Design (Vibrations) from Delhi Technological University, Delhi, India. Presently, she is working as an Assistant Professor in the Mechanical, Production, Industrial and Automobile Engineering Department at Delhi Technological University, Delhi, India. Her research interest is on fault diagnosis in Machinery.

FPGA-Based Acceleration of a Neural Network for Symbol Classification with Hardware-Optimised Multiplier Implementation

Lakshay Chandna
Dept. of Electrical Engineering
Delhi Technological University (DTU)
Delhi, India
itslakshay179@gmail.com

Megha Sharma
Dept. of Electrical Engineering
Delhi Technological University (DTU)
Delhi, India
megha8747@gmail.com

Kushik Agarwal
Dept. of Electrical Engineering
Delhi Technological University (DTU)
Delhi, India
agarwal.kushik789@gmail.com

Anup Kumar Mandpura
Dept. of Electrical Engineering
Delhi Technological University (DTU)
Delhi, India
kanup@dtu.ac.in

Himanshu Singh
Dept. of Electrical Engineering
Delhi Technological University (DTU)
Delhi, India
himanshusingh@dtu.ac.in

2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe) | 979-8-3503-4743-2/23/$31.00 ©2023 IEEE |
DOI: 10.1109/EEEIC/ICPSEUROPE57605.2023.10194730

***Abstract*— Symbol classification is a crucial task in many applications such as handwriting recognition, computer vision, Optical Character Recognition (OCR), etc. However, multiclass classification is an intensive task that requires high computational power and memory resources. In this paper Field-Programmable Gate Array (FPGA) based acceleration of a neural network for symbol classification has been implemented with hardware-optimised multiplication techniques. The approach involves using FPGA to perform faster multiplications and accumulation operations with reduced resource usage. The neural network is trained using MATLAB and then implemented in VHDL on Quartus Prime Lite. The multiplication is compared to the Karatsuba Multiplication technique and the Iterative Logarithmic technique. Based on the comparative study the Iterative Logarithmic technique proved to be the best to bring down the overall hardware usage. The final optimised design has been validated using ModelSim and then deployed on an Altera Cyclone® V SE FPGA device.**

Keywords — Symbol Classification, Neural Networks, FPGA, Karatsuba Multiplier, Iterative Logarithmic Multiplier

I. INTRODUCTION

Big Data, Machine Learning (ML), the Internet of Things (IoT) are among the data-intensive applications that are driving the demand for accelerators to offload work from general-purpose processors. There are numerous accelerators available, including Application-Specific Integrated Circuit (ASIC), Graphics Processing Unit (GPU), and Field Programmable Gate Array (FPGA) devices.

Dedicated hardware acceleration can be implemented with high performance and flexibly with FPGAs. FPGA accelerators have become a popular choice in data centres over the past ten years because of their adaptability and programmable nature. These devices are a good solution for applications requiring real-time output and high portability, but due to their high cost, one must also consider optimising the design for deployment on relatively low-end hardware by implementing alternate modern techniques in accordance with the available resources at a lower cost.

FPGA-based acceleration of Neural Networks (NNs) has been a prevalent research topic in the past couple of years. In [1] the authors proposed a robust and re-configurable architecture for deploying ML algorithms. The paper implemented the architecture to design, train, and test the Back Propagation Neural Network using FPGA. The results showed that using an FPGA is faster but slightly less accurate. The framework implemented in [2] makes NN computable in FPGAs. The proposed framework reduced the required resources and increased the performance with different approximation techniques. In [3] a general NN algorithm was implemented on the XILINX ZU9CG FPGA SOC platform which showed an outperformance in energy efficiency and real-time computation.

FPGA-based acceleration is an effective and efficient solution for solving the Artificial Intelligence (AI) and ML problems of the real–world. In [4] NN was created to identify and classify handwritten digits using the Modified National Institutes of Standards and Technology (MNIST) dataset. The network was trained and tested on this dataset. The hardware implementation of the network was accomplished through the use of Vivado High-Level Synthesis (HLS) software, which involved programming in Verilog. In reference [5], a neural network was used to recognize handwritten digits. The network was tested using 100 samples from the MNIST database, and it was simulated on the Quartus II 13.0 and Modelsim FPGA development platforms. Using hardware-based acceleration gives a huge enhancement in computing speed but this happens by using more complex and expensive digital designs of FPGA devices with sufficient resources [6].

In [7] 70% accuracy was achieved for the forest fire detection method through NN by processing the data from several sensors. A low-cost Convolutional Neural Network (CNN) based real-time Traffic Sign Recognition hardware accelerator was proposed in [8] by the authors, which enables the vehicle to recognize road signs using a camera and assist the driver. In [9] authors have implemented a single switch

979-8-3503-4743-2/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:23:09 UTC from IEEE Xplore. Restrictions apply.

Jerk chaotic oscillator prediction model based on Feed Forward NN on Intel Cyclone IV.

One of the main challenges in implementing neural networks on FPGAs is reducing computational complexity while maintaining accuracy. The basic arithmetic units required for a neuron unit are adders and multipliers. While addition can be implemented with very low complexity, multiplication is complex. Efficient architectures require low complexity and fast multipliers.

Some of the papers have proposed hardware-optimised algorithms like Karatsuba, Dadda, Iterative Logarithm, etc. In [10], the author proposed an FPGA-based modified Karatsuba Multiplier which was 26.5% faster than the traditional Karatsuba multiplier. The authors in [11] proposed a new method to improve the accuracy and performance of iterative logarithmic multiplier based on Mitchell's algorithm. The authors in [12] claim that the hardware implementation of an iterative logarithmic multiplier uses only one adder and some shifters, which results in reduced usage of total logic resources.

To the best of our knowledge, the implementation of the Neural Network based symbol classifier using an iterative logarithmic multiplier for FPGA-based hardware acceleration has not been implemented yet. In this paper, we propose FPGA-based acceleration of symbol classification Neural Network and compare different multiplier implementation techniques for FPGA-based applications, based on their hardware usage while preserving the accuracy of the model.

II. Problem Statement

We consider an input image of size 1280x720 pixels comprising multiple instances of four types of symbols. The NN Model designed would be able to identify the four symbols as shown in Fig. 1 and label them by assigning them different colours.

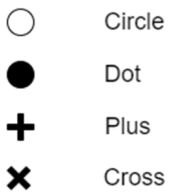


Fig. 1. Input Symbols

As the multiplication was resource-intensive, alternative multiplying algorithms were explored. The most effective multiplier technique is then employed to deploy NN over FPGA based on the comparative study of multipliers.

III. Methodology

The overall system involves various stages as shown in Fig. 2. The implementation of the project starts with training of the NN using MATLAB for obtaining the weights and biases. These are then used for the VHDL implementation that can be deployed on an FPGA board, after optimising the circuit design.

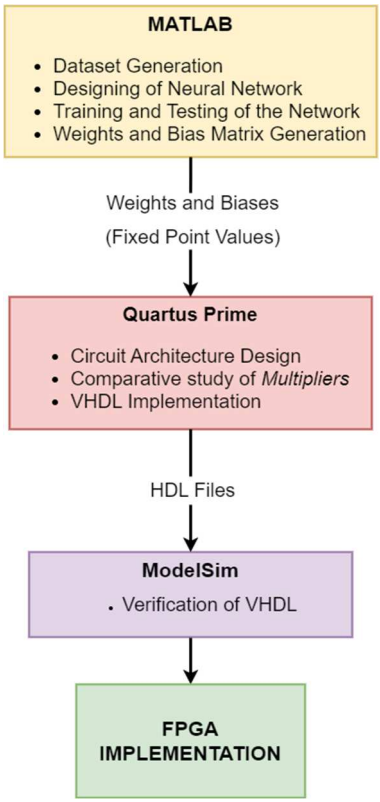


Fig. 2. System Model

A. Generation of Training Dataset

An artificial NN as shown in Fig. 3 is trained using input images that have a size of 7x7 pixels. Every image is transformed into a matrix of pixel values. Each pixel's value ranges from 0 (black) to 255 (white). In order to expand the training dataset, pixel shifting has been performed on all training images to cover cases where only partial symbols are present in the input image region. This resulted in the expansion of the dataset to 200,000 images.

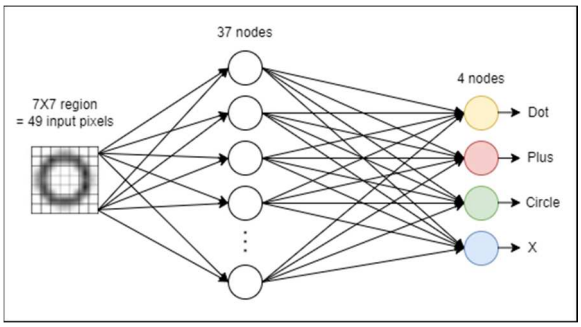


Fig. 3. Neural Network

B. Designing, Training & Testing of NN

A fully connected feed-forward CNN was designed and trained on MATLAB with one hidden layer containing 37 nodes (Fig.3). Sigmoid function is used as the activation function in the CNN.

The parameters of the Neural Network used are listed in Table I.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:23:09 UTC from IEEE Xplore. Restrictions apply.

TABLE I. NEURAL NETWORK PARAMETERS

S No.	Neural Network Parameters	Values
1.	Number of layers	3 (Input, 1 hidden, output layer)
2.	Input Layer Nodes	49 nodes
3.	Hidden layer	37 nodes
4.	Output layer	4 nodes
5.	Epochs	400
6.	Learning Rate	0.00001
7.	Activation Function	Sigmoid

Here, 37 nodes in the hidden layer are calculated via the Rule of Thumb (⅔ of input nodes plus output nodes). Four nodes in the output layer correspond to the four symbols.

C. Circuit Architecture

Once the training is performed on MATLAB, two matrices are generated. These matrices consisted of weights and biases for input to the hidden layer and for hidden layer to the output layer. Both of these matrices are converted to a fixed-point representation and then a VHDL file is generated for the FPGA implementation.

The NN is implemented in VHDL using Quartus Prime Lite. The input image can be RGB which is converted to grayscale using the luminance technique, with each pixel value being represented by an 8-bit number (0-255). The grayscale image is stored using a memory unit. The input image is divided into portions of 7x7 pixels and traversed using registers. Once the 1280th pixel is reached, shift a line vertically until all 720 lines are covered as the size of the input test image taken is 1280x720 pixels.

The circuit architecture for the first layer with 49 input signals is shown in Fig. 4. The 7x7 section of the image is stored and given as input to the neural network using seven line memories (LM) having six registers (R) each, for introducing appropriate delay, and one direct connection. Then all 49 input values are fed into the 37 hidden layer neurons.

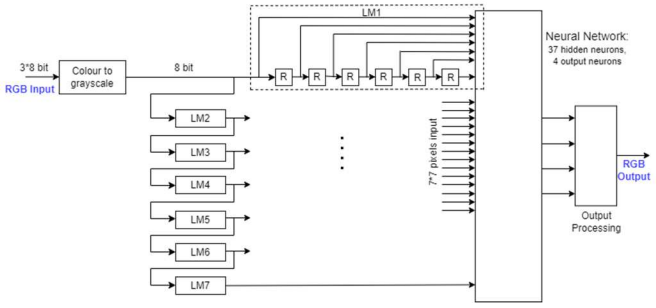


Fig. 4. Circuit Architecture Design

Fig. 5 shows the neuron structure where we get the input signals (x1, x2, x3), then multiply them with their respective weights (w1, w2, w3), add the bias, and send it to the sigmoid function (in ROM) that gives the output value in the range of 0 to 1. The model described in this paper gives out four values for each input section of 7x7. These four values correspond to the four different symbols, only one of these values is greater than 0.5 which helps the model to declare the final classification result.

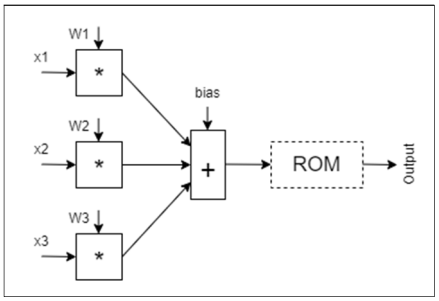


Fig. 5. Neuron Structure

D. Multiplier Techniques

The traditional multiplier is implemented using DSP blocks, if available, but in ML applications the number of multipliers required is high so to optimise this Karatsuba Multiplier and Iterative Logarithmic Multiplier have been implemented, and compared in terms of hardware usage. The algorithm behind these multipliers has been discussed below -

1. *Karatsuba Multiplier:*

The Karatsuba multiplier is an algorithm used for the fast multiplication of large numbers. The basic idea behind the Karatsuba multiplier is to divide a large number into smaller parts and perform some simple arithmetic operations mentioned below on these parts and then recombine the results to get the final answer.

The multiplication of two integers *'x'* and *'y'* using Karatsuba multiplier can be done as follows:

$$Divide\ x\ into\ x_H,\ x_L$$
$$Such\ that\ x = x_H z^{\frac{n}{2}} + x_L$$
$$where\ z = 10\ (in\ case\ of\ integers)$$
$$and\ n = Number\ of\ Digits$$
$$Similarly, y\ can\ be\ divided\ into\ y_H,\ y_L$$
$$Let \quad D_0 = y_L x_L \quad (1)$$
$$D_1 = y_H x_H \quad (2)$$
$$D_2 = (y_L + y_H)(x_L + x_H) - y_L x_L - y_H x_H \quad (3)$$
$$p = D_1 z^n + D_2 z^{n/2} + D_0 \quad (4)$$

where p is the final product

Time complexity of Karatsuba multiplier:

$$O(n^{log_2 3}) = O(n^{1.59})$$

Time complexity of traditional multiplier:

$$O(n^2)$$

where n is the number of digits in the two numbers being multiplied.

2. *Iterative Logarithmic Multiplier:*

Mitchell algorithm-based iterative logarithmic multiplier is a type of digital logic circuit that can iteratively perform the operation of multiplication, using logarithmic processing techniques. The basic idea behind this circuit is to perform the multiplication of two numbers by successively shifting and adding, rather than using traditional methods such as a lookup table or an array of full adders. The multiplication method for two integers N_1 *and* N_2 is as follows:

$$Product\ (P) = N_1 N_2$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:23:09 UTC from IEEE Xplore. Restrictions apply.

The binary representation of any integer number:

$$N = 2^k(1+x) \quad (5)$$

where x is the fractional part.

$$log_2 N = k + log_2(1+x) \quad (6)$$

The logarithm of the multiplication of N_1 and N_2:

$$log_2(N_1N_2) = log_2(N_1) + log_2(N_2) = k_1 + log_2(1+x_1) + k_2 + log_2(1+x_2) \quad (7)$$

The expression $log_2(1+x)$ is approximated by x and the logarithm of the product of two numbers is expressed as the sum of their characteristic numbers and mantissa:

$$log_2(N_1N_2) = k_1 + x_1 + k_2 + x_2 \quad (8)$$

$$N_1N_2 = 2^{k1+k2}(x_1+x_2) \quad when\ x_1+x_2 >= 1$$
$$N_1N_2 = 2^{k1+k2}(1+x_1+x_2) \quad when\ x_1+x_2 < 1 \quad (9)$$

The proposed MA-based multiplication produces a significant error percentage. The authors in [12] proposed an iterative algorithm to reduce the error as much as possible and obtain accurate results. From the binary representation of numbers in (5), we can derive a correct expression for multiplication (P_{true}):

$$P_{true} = N_1N_2 = 2^{k1}(1+x_1).\ 2^{k2}(1+x_2) = 2^{k1+k2}(1+x_1+x_2)+2^{k1+k2}(x_1.x_2) \quad (10)$$

To avoid the approximation error, we have to take into account the next relation derived from (5):

$$x.2^k = N - 2^k \quad (11)$$

Therefore,

$$P_{true} = N_1N_2 = 2^{k1+k2} + (N_1 - 2^{k1})2^{k2} + N_2.2^{k1} - 2^{k2+k1} + (N_1 - 2^{k1}).(N_2 - 2^{k2}) \quad (12)$$

Let,

$$P^{(0)}_{approx} = 2^{k1+k2} + (N_1 - 2^{k1})2^{k2} + (N_2 - 2^{k2})2^{k1} \quad (13)$$

be the first approximation of the product. It is evident that

$$P_{true} = P^{(0)}_{approx} + (N_1 - 2^{k1}).(N_2 - 2^{k2}) \quad (14)$$

We can calculate $P^{(0)}_{approx}$ and repeat the procedure until the exact result is obtained.

The absolute error after the first approximation is:

$$E^{(0)} = P_{true} - P^{(0)}_{approx} = (N_1 - 2^{k1}).(N_2 - 2^{k2})$$
$$E^{(0)} = C^{(1)} + E^{(1)} \quad (15)$$

where C(1) is the approximate value of E(0) and E(1) is an absolute error when approximating E(0).

$$P_{true} = P^{(0)}_{approx} + C^{(1)} + E^{(1)} \quad (16)$$

We can now add the E(0) estimate to the product approximate P_{approx} as a correction term with which we can reduce the error in approximation.

$$P^{(1)}{}_{approx} = P^{(0)}{}_{approx} + C^{(1)} \quad (17)$$

By repeating this multiplication procedure for i correction terms, the product can be approximated as:

$$P^{(i)}{}_{approx} = P^{(0)}{}_{approx} + C^{(1)} + C^{(2)} + \ldots + C^{(i)} = P^{(0)}{}_{approx} + \sum_{j=1}^{i} C^{(j)} \quad (18)$$

Time complexity of iterative logarithmic multiplier:

$$O(log(n))$$

Time complexity of traditional multiplier:

$$O(n^2)$$

where, n is the number of digits in the two numbers being multiplied

Hence, different multiplication algorithms were compared so that the designed model can fit on the selected FPGA board, i.e. *DE0-Nano-SoC (Altera Cyclone® V SE 5CSEMA4U23C6N device).*

IV. Results

Accuracy and Cost were calculated and graphs were plotted after the NN training was completed in MATLAB. Accuracy calculation was done by calculating the fraction of correct predictions upon total predictions. In Fig. 6 we can see that the accuracy gradually increases with increasing epochs and settles around 96.3%.

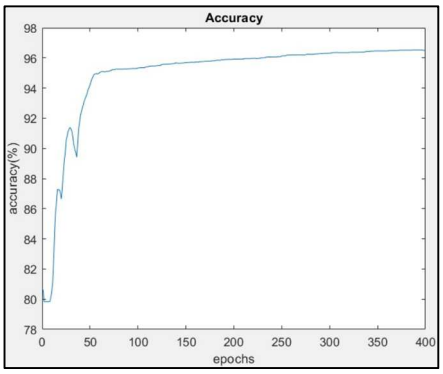


Fig. 6. Accuracy Graph

The cost function calculates the mean squared error in the predictions made.

$$L_{square}(\hat{y}, y) = (\hat{y} - y)^2 \quad (19)$$

$$Cost(f, D) = \frac{1}{n}\sum_{i=1}^{n} L(\hat{y}_i, y_i) \quad (20)$$

where, L_{square} = Loss function

$\hat{y}_i$ = Predicted value

y_i = Actual value

When the cost is plotted against epochs in Fig. 7, it can be observed that the cost function comes down to 0.042, which is an acceptable value.

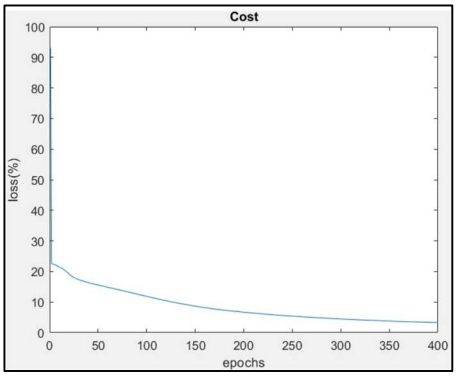


Fig. 7. Cost Graph

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:23:09 UTC from IEEE Xplore. Restrictions apply.

As mentioned earlier, all three multiplication techniques were implemented and compiled, and the results have been reported in TABLE II. The number of available resources in the selected FPGA Board were:
Adaptive Logic Modules (ALMs)= 15,880
Total Digital Signal Processing (DSP) Blocks = 84

Further fitter analysis was done to make sure the target device's resources could support the logic and time requirements of the design. Only one of the three techniques i.e. Iterative Logarithmic multiplier, was able to successfully fit on our chosen FPGA board, as shown in Table II.

TABLE II. COMPARATIVE STUDY

MULTIPLIER TECHNIQUE	Resources Required			Fitter Result
	ALMs	*Total Registers*	*DSP Blocks*	
Traditional	12344	11288	252	Unsuccessful
Karatsuba	28201	13953	235	Unsuccessful
Iterative Logarithmic	9336	12430	0	Successful

From the study, we can observe that an iterative logarithmic multiplier can be used for reducing the number of hardware resources required, especially when there are a few or no DSP blocks present. As a result of this reduction in resource usage, the model was able to successfully fit on the selected Cyclone V FPGA board. Earlier the model was able to do so just for Cyclone 10 series boards, which are highly expensive and advanced. Once the fitter analysis was complete for a synthesised design, Modelsim Simulation was launched using the testbench, with an input test image (1280x720 pixels) of which a small section is shown in Fig.8.The expected Classification Results (colour coded):

- Circle => Green
- Dot => Yellow
- Plus => Red
- X => Blue

Output image of the section shown in Fig. 8 can be seen in the Fig. 9.

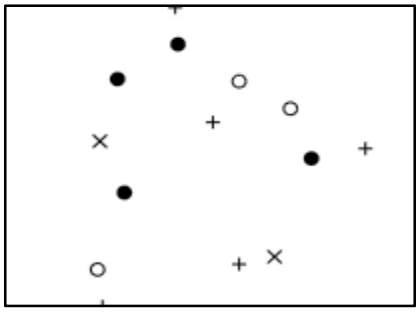


Fig. 8. Section of Input Image

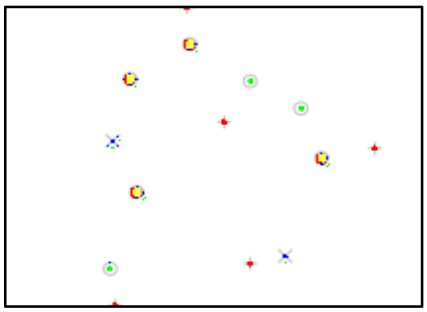


Fig. 9. Section of Output Image

V. CONCLUSION

This paper has presented an FPGA-based acceleration of NN for symbol classification with hardware-optimised multiplier implementations. After successful implementation and comparison of multipliers, it was found that the Iterative Logarithmic multiplier gives the best results as it uses equivalent ALMs compared to the default but it doesn't use any DSP Blocks, which helps in deploying the model on an FPGA board with few available resources. The experimental results show that the proposed approach significantly enhances the speed and efficiency of symbol classification tasks, making it suitable for real-time applications with limited computational resources.

REFERENCES

[1] A. Itagi, S. Krishvadana, K. P. Bharath and M. Rajesh Kumar, "FPGA Architecture To Enhance Hardware Acceleration for Machine Learning Applications," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1716-1722, doi: 10.1109/ICCMC51019.2021.9418015.

[2] S. T. Brassai, A. Hammas and B. Bustya, "Framework for neural network hardware implementation," 2022 23rd International Carpathian Control Conference (ICCC), Sinaia, Romania, 2022, pp. 387-391, doi: 10.1109/ICCC54292.2022.9805981.

[3] Hao, Y. (2017). A General Neural Network Hardware Architecture on FPGA. ArXiv. /abs/1711.05860

[4] S. S. Lingala, S. Bedekar, P. Tyagi, P. Saha and P. Shahane, "FPGA Based Implementation of Neural Network," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ACCAI53970.2022.9752656.

[5] L. zenggang, W. zhengyan and S. jingcheng, "Design of Neural Network Identification System Based on FPGA," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 2369-2372, doi: 10.1109/IAEAC47372.2019.8997779.

[6] S. Ivanov, S. Stankov and T. Nenov, "FPGA Based Neural Networks for Characters Recognition," 2018 20th International Symposium on Electrical Apparatus and Technologies (SIELA), Bourgas, Bulgaria, 2018, pp. 1-3, doi: 10.1109/SIELA.2018.8446658.

[7] S. Anand and Keetha Manjari.R.K, "FPGA implementation of artificial Neural Network for forest fire detection in wireless Sensor Network," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 265-270, doi: 10.1109/ICCCT2.2017.7972284.

[8] J. Kim, J. -K. Kang and Y. Kim, "A Low-Cost Fully Integer-Based CNN Accelerator on FPGA for Real-Time Traffic Sign Recognition," in IEEE Access, vol. 10, pp. 84626-84634, 2022, doi: 10.1109/ACCESS.2022.3197906.

[9] M. AlFarah and T. Bonny, "Chaotic Oscillator Prediction Based on Artificial Neural Network and its Realisation on FPGA," 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 2022, pp. 1-4, doi: 10.1109/ASET53988.2022.9735071.

[10] G Samanta, Jagannath, Razia Sultana, and Jaydeb Bhaumik. "FPGA based modified Karatsuba multiplier." In Proc. Int. Conf. VLSI Signal Process.(ICVSP), vol. 10, p. 12. 2014.

[11] Z. Babić, A. Avramović, P. Bulić, "An iterative logarithmic multiplier" Microprocessors and Microsystems, Volume 35, Issue 1,2011,Pages 23-33,ISSN 0141-9331.

[12] Lotrič, U., Bulić, P. (2011). Logarithmic Multiplier in Hardware Implementation of Neural Networks. In: Dobnikar, A., Lotrič, U., Šter, B. (eds) Adaptive and Natural Computing Algorithms. ICANNGA 2011. Lecture Notes in Computer Science, vol 6593. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20282-7_17

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:23:09 UTC from IEEE Xplore. Restrictions apply.

Nanotechnology

ACCEPTED MANUSCRIPT

Graphene oxide/Mn3O4 nanocomposites for advanced electrochemical biosensor for fumonisin B1 detection

To cite this article before publication: Sandeep Sarpal *et al* 2023 *Nanotechnology* in press https://doi.org/10.1088/1361-6528/aceaff

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is **© 2023 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript will be available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.98 on 10/08/2023 at 07:57

Graphene oxide/Mn_3O_4 Nanocomposites for Advanced Electrochemical Biosensor for Fumonisin B1 Detection

Sandeep Sarpal[a,b], Avinash Kumar Singh[b], Hema Bhardwaj[b], Nitin Kumar Puri[a]*, Pratima R. Solanki[b]*

[a]Department of Applied Physics, Delhi Technological University, New Delhi, India

[b]Special Centre for Nanoscience, Jawaharlal Nehru University, New Delhi, India

***Corresponding Author**: partima@mail.jnu.ac.in; nitinkumarpuri@dtu.ac.in

Abstract

Occurrence of mycotoxins in food samples threat to its safety issue due to the presence of high toxicity and carcinogenic behavior, thus requiring highly sensitive and selective detection. Herein, the trimanganese tetraoxide (Mn_3O_4) nanoparticles in combination with graphene oxide (GO) nanocomposite were used to enhance the electrochemical performance for fabrication of electrochemical biosensor for fumonisin B1 (FB1) detection. The various characterization tools were used to validate the fabrication of GOMn_3O_4 nanocomposites. To fabricate the electrochemical biosensor on an indium tin oxide (ITO) coated glass substrate, a thin film of GOMn_3O_4 nanocomposite was prepared using electrophoretic deposition (EPD) technique, and antibodies (ab-FB1) were immobilized onto the electrode for selective FB1 detection. The differential pulse voltammetry (DPV) technique was used to observe the sensing performance. The non-binding sites of the ab-FB1 on the immunoelectrode were blocked with bovine serum albumin (BSA). The biosensor electrode was fabricated as BSA/ab-FB1/GOMn_3O_4/ITO for the detection of FB1. The sensitivity of the biosensor was obtained as 10.08 μA mL ng^{-1} cm^{-2} in the detection range of 1 pg mL^{-1} to 800 ng mL^{-1} with a limit of detection (LOD) of 0.195 pg mL^{-1}. In addition, the recovery of BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrodes was also performed on sweet corn samples and is calculated to be 98.91%.

Keywords: Biosensor; graphene oxide; nanocomposite; immunoelectrode; and fumonisin B1.

Accepted Manuscript

1. Introduction

The occurrence of mycotoxins in food items is a global challenge for food contamination, and it is considered as important toxin that affect onto human and animal health. Mycotoxins are naturally occurring organic toxic compounds that are produced by secondary metabolites of fungi such as *Aspergillus*, *Penicillium*, *Fusarium*, *Claviceps,* etc. in food and feed products [1,2]. Annually, 25% of the agricultural harvested crops are contaminated by mycotoxin and causing huge economic loss and responsible for several acute and chronic diseases worldwide. Among the presence of various mycotoxins, fumonisin B1 (FB1) is one of important mycotoxin that predominates in 70% of total FBs, naturally produced by fungi *Fusarium verticillioides* and *Fusarium proliferatum* are occurred in many food and feed products e.g. peanuts, maize-based products, and cereal-based products, etc. [3,4]. FB1 has also been designated as a carcinogen by the International Agency for Research on Cancer (IARC), placing it in the class 2B group of human carcinogens. The toxicity of FB1 causes several health issues such as brain damage, lung, and kidney failure, etc. [5]. Thus, the research on the detection of FB1 has become very important.

There are several analytical techniques available for detecting FB1 mycotoxin. Mainly these techniques belong to conventional tools such as gas Chromatography-Mass Spectrometry (GCMS) [5], enzyme-linked immunosorbent assay (ELISA) [6], high-performance liquid chromatography (HPLC) [7], liquid chromatography-mass spectrometry (LCMS) [8] and polymerase chain reaction (PCR) [9], etc. These techniques are sensitive, specific, and selective toward their analytes. However, these techniques have several disadvantages such as being sophisticated, requiring more samples for detection, very tedious processing, and more importantly, requiring a well-trained person to operate the system. To overcome the aforementioned difficulties, an electrochemical biosensor is a very promising tool for detecting analytes at very low concentrations and has several advantages over conventional techniques.

Nanomaterials have been considered a great interest in the application of sensors or biosensors development. Among a variety of nanomaterials, graphene-based nanomaterials have emerged as the most advanced materials for sensing applications

Accepted Manuscript

due to their enhanced signal response. Graphene-based nanomaterials provide a large surface area and excellent biocompatibility with a wide range of biomolecules such as proteins, cells, DNA, enzymes, and antibodies [9]. Graphene also has high electrical conductivity, high electron transfer rates, and the ability to immobilize various molecules [10]. Graphene-based biosensors also offer extremely high sensitivity, stability, and fast response times, leading to advances in healthcare and diagnostics.

Manganese-based materials have unique redox properties which give significant importance in the applications related to energy, environment, and biomedical fields [11] Mn_3O_4 nanomaterial is also known as the oxidizing and reducing agent, and it is a spinel-structured stable oxide with unique catalytic, electrochemical, and magnetic properties. It is a p-type semiconductor having a larger bandgap of 2.3 eV which provides supercapacitor properties to it [12]. It has several physical properties like thermodynamic stability, high electrochemical conductivity, chemical stability, paramagnetic behaviour, etc [13,14]. Mn_3O_4 is the unique phase of manganese oxide nanomaterial which is explored in various applications such as excellent supercapacitors, high-density magnetic storage devices, high charge carrier mobility, good biocompatibility, etc [14,15]. Moreover, Mn_3O_4 nanomaterial exhibits distinguished properties like environmental compatibility and cost-effectiveness. However, the Mn_3O_4 phase of manganese oxide is biologically less explored. Graphene oxide (GO) sheet has more functional groups on its basal plane and edges, and it results in the enhancement of the loading amount of Mn_3O_4, surface area, and porosity [16]. GOMn_3O_4 nanocomposite exhibits higher energy density and specific capacitance as compared to GO, while larger specific capacitance and power density as compared to Mn_3O_4. The nanocomposites of nitrogen doped GO and manganese oxide for photodynamic therapy magnetic resonance imaging of cancer cells without affecting the normal tissue. Their results of cell viability analysis showed the significant cell death under laser irradiation. The elemental composition of GOMn_3O_4 nanocomposites showed intense signals which showed that the 'Mn' was the major element showing the optical absorption [17]. Furthermore, Chen et al. developed nanocomposites of reduced graphene oxide wrapped Mn_3O_4 as anode materials for high-performance lithium ion batteries which showed the specific capacity of 795.5 mA hg^{-1} after 200 cycles. The

Accepted Manuscript

nanocomposites showed the excellent cyclic performance and good electrochemical activity [18].

Several research groups have reported various methods for detecting mycotoxins. For instance, Ahmed et al recently developed an electrochemical biosensor for detecting the naturally occurring histamine based on the molecularly imprinted polymer using the gold inverse opal electrode and the detection limit of 1.07 μM [19]. Dhiman et al developed the silver-cerium oxide (Ag-CeO_2) nanocomposite-based electrochemical biosensor for FB1 detection [20]. Furthermore, Singh et al developed a molecularly imprinted polymer-based electrochemical sensing tool for FB1 detection [3]. Alsulami et al developed a homogeneous luminescence immunoassay for FB1 detection [21]. Zhang et al developed graphene and gold nanocomposites based on label-free DNA for the detection of FB1 [22]. In recent years, Mao et al developed GO-CdS heterojunctions based on an ultrasensitive molecularly imprinted photoelectrochemical sensing platform for detecting FB1 [23]. Wu et al developed the fluorescence resonance energy transfer (FRET) system for the detection of FB1 [24]. Jodra et al developed the electrochemical magneto immunosensor involving magnetic beads for the detection of fumonisins [25]. For the detection of FB1, Lu et al created a competitive fluorescence enzyme-linked immunosorbent test based on the quenching of cadmium telluride (CdTe) quantum dots [26]. Yang et al created an electrochemical immunosensor using single-walled carbon nanotubes and chitosan to detect FB1 in corn samples [27]. Chen et al created a colloidal gold-based lateral flow immunoassay using modified gold nanobeads to detect FB1 [28]. Furthermore, Qu et al created a magnetic beads-assisted fluorescence aptasensing method to detect FB1 [29]. Table 1 shows comparison data between current work and previously reported results for FB1 detection.

Herein, simple and sensitive electrochemical immunosensor was fabricated for the detection of FB1 mycotoxins using GOMn_3O_4 nanocomposite. This nanocomposite was synthesized at room temperature which is robust and cost-effective. The phase and crystallinity of the synthesized GOMn_3O_4 nanocomposite were observed by XRD and Raman spectroscopy. The immunosensor was fabricated onto the ITO glass substrate using a GOMn_3O_4 film by an EPD method, which was used for the immobilization of

Accepted Manuscript

antibodies (ab-FB_1). The non-specific binding sites of antibodies were blocked by the BSA protein. The fabricated immunosensor (BSA/ab-FB1/GOMn_3O_4/ITO) showed a wide detection range as a function of FB1. Here, results suggest that the newly designed immunosensor is robust, cost-effective, sensitive, selective, and selective towards FB1.

2. Experimental Section

2.1 Material used

Graphite powder (20-micron), hydrochloric acid (HCl), orthophosphoric acid (H_3PO_4), potassium permanganate ($KMnO_4$), sulfuric acid (H_2SO_4) FB1-specific ab-FB_1, ochratoxin A (OTA), deoxynivalenol (DON), cellulose, fructose, starch, FB1, manganese (II) acetate tetrahydrate ($(CH_3COO)_2Mn.4H_2O$), sodium hydroxide (NaOH) was purchased from M/S Sigma Aldrich. The glass substrate coated with indium tin oxide (ITO) coated glass substrate was used with a sheet resistance of 10 Ω sq^{-1} having a thickness of 1.1 mm. Sodium phosphate monobasic anhydrous (NaH_2PO_4) and sodium phosphate dibasic dehydrate (Na_2HPO_4), N-ethyl-N-(3-dimethylamino propyl) carbodiimide (EDC), N-hydroxysuccinimide (NHS), sodium chloride, potassium chloride, BSA were purchased from SRL, India. Potassium ferricyanide [$K_3(Fe(CN)_6)$] and potassium ferrocyanide [$K_4(Fe(CN)_6)$] were obtained from Fischer Scientific. Deionized water resistivity of 18.2 MΩ.cm was purified by Millipore water purification assembly and used for the preparation of buffer and all other required solutions. Phosphate buffered solution (PBS, 0.2 M) was prepared containing 0.9% NaCl and 5 mM of redox probe $[Fe(CN)_6]^{3-/4-}$ with pH 7.0

2.2 Synthesis of GO

GO was synthesized by a modified Hummer's method using pure graphite powder (20 μm) [30]. In the typical procedure, 240 mL of H_2SO_4 and 27 mL H_3PO_4 were mixed in a beaker under continuous magnetic stirring for 30 minutes. Then, 12 gram of $KMnO_4$ was added very slowly in small steps. The mixture was stirred using a magnetic stirrer for 15 hours until the solution became dark green. After this, 2 mL of H_2O_2 was slowly added dropwise and stirred for 30 minutes. The ice was placed outside to prevent heating due to the occurrence of an exothermic reaction. Then, it was allowed to cool down. After this, the mixture was washed several times with (70%) distilled water and

Accepted Manuscript

(30%) HCl and centrifuged using Remi Make Eppendorf centrifuge at 400 rpm for 10 minutes. Then supernatant was decanted away and was rewashed again and again until the pH reached to 7. The washed GO solution was dried using a vacuum oven at 60 ^{o}C for 24 hours and powder of GO was collected.

Synthesis of Mn_3O_4 nanoparticles

The Mn_3O_4 nps were synthesized using the co-precipitation method [31]. In the typical synthesis, 0.2 M of ($(CH_3COO)_2Mn.4H_2O$) and 0.5 M of NaOH were used as precursors. NaOH was added dropwise to manganese acetate using constant magnetic stirring of 800 rpm at room temperature until the pH of the solution reached~12. Then the precipitate was filtered out and washed many times using the distilled water until the pH of the solution reached 7. The obtained precipitate was then dried at room temperature using the vacuum oven.

Synthesis of GOMn_3O_4 nanocomposite

The GOMn_3O_4 nanocomposite was synthesized using the one-step *in-situ* preparation process [32]. GOMn_3O_4 nanocomposite was prepared in the presence of GO. In the typical synthesis, 200 mg of GO was mixed in 100 mL of distilled water at room temperature and sonicated. Then dispersed GO was added slowly to the solution containing 0.2 M manganese acetate, and 0.5 M reducing agent (NaOH) until the completion of the reaction. The obtained pH of the solution reached~12. Then the precipitate was filtered out and washed several times until the value of pH reached to 7. The obtained precipitate was then dried at room temperature using the oven and collected in a vial GOMn_3O_4 nanocomposite.

2.3 Characterization Techniques

DC power supply of 60V was used to fabricate thin film of nanocomposite on the ITO glass substrate. The crystallinity of GOMn_3O_4 nanocomposite was investigated using an X-ray diffraction (XRD) spectrometer (PANalytical Model X'pert Pro with Cu K_α radiation at λ = 1.5406 Å). The working voltage was 45 kV and the current was 40 mA. The XRD patterns were recorded in the range of 10^{o} to 80^{o} with a step size of 0.033^{o} and

Accepted Manuscript

a scan rate of 2° per minute. The Raman spectra were captured using the Alfa 300 witec Raman spectrometer with 532 nm laser. A JEOL 2100F High Resolution Transmission Electron Microscopy (HR-TEM) with a 200 KV accelerating potential was used to capture images. The Carl-Zeiss EVO40 was used to study the energy dispersive X-ray (EDX) spectra of the nanocomposite. Nova nano FE-SEM-450 was used to capture the Field-emission scanning electron microscopy images (FE-SEM) of the graphene, and the nanocomposite. Fourier Transform Infra-red (FTIR, UATR TWO) spectroscopy of Perkin Elmer was used for the investigation of the vibrational spectra of graphene, and the nanocomposite. The electrochemical analyzer (Autolab-AUT-Metrohm) was used to investigate the electrochemical characteristics (AUT86014) in a three electrode system composed of ITO as a working electrode, Platinum as a counter and Ag/AgCl as a reference electrode PBS containing 5 mM $[Fe(CN)_6]^{3-/4-}$.

2.4 Thin film fabrication of $GOMn_3O_4$/ITO electrode using EPD Technique

Glass surface coated with ITO electrode were first hydrolyzed using the mixture of H_2O:NH_3:H_2O_2 in the ratio of 5:1:1 to enhance the hydrophilic nature of ITO coated glass and kept in the oven for about 1 hour at 80°C. Then the ITO coated glass substrates were washed with a mixture of ultra-pure water and ethanol and dried at 60 °C for about 1 hour. These hydrolyzed ITO-coated glass substrates were further used for the fabrication of immunosensor for FB1 detection. For this, synthesized $GOMn_3O_4$ nanocomposite was deposited onto the ITO coated electrode using EPD technique. The EPD method constitutes an economical, eco-friendly, low-energy-consuming alternative to the high energy consuming evaporative techniques [33]. To obtain uniform film of $GOMn_3O_4$ onto ITO surface**,** 5 mL of $GOMn_3O_4$ solution was mixed in a 3 mL acetonitrile solution then deposited onto ITO surface by applying an optimized D.C. 60 volt for 5 minutes to obtain smooth and uniform thin films.

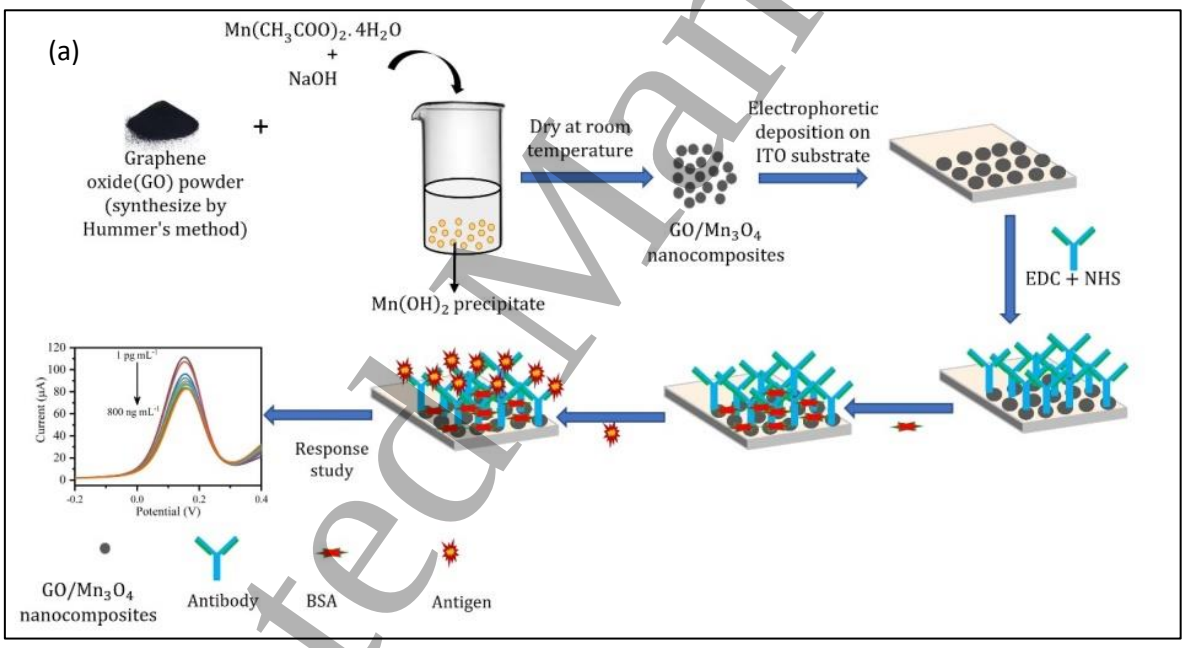
2.5 The fabrication process of Immunoelectrode

A solution of 5% ab-FB1 (in sodium phosphate buffer), EDC (0.4 M), and NHS (0.1 M) were synthesized in a 2:1:1 ratio and placed at 4 °C for 1 hour. For the immobilization process, 20 mL of a solution of activated EDC-NHS was evenly dropped onto the surface of the $GOMn_3O_4$/ITO electrode and placed at room temperature in a humidity

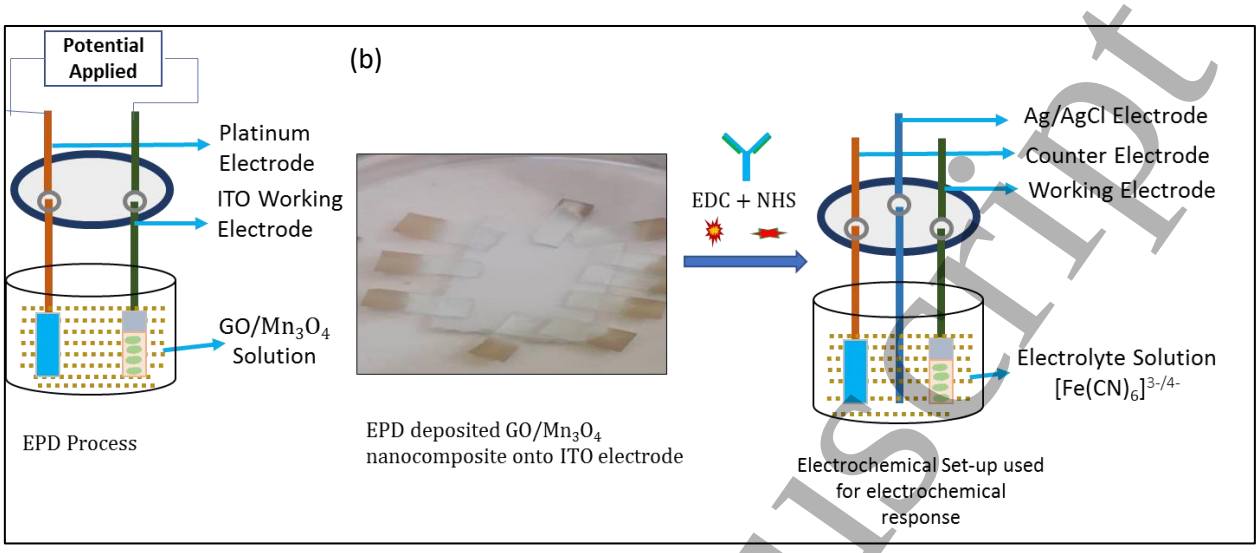
Accepted Manuscript

chamber for 6 hours. The positively charged GOMn_3O_4 nanocomposite and the carbonyl group of ab-FB1 create an electrostatic interaction that leads to the formation of metal ester bonds.

PBS was used to clean the ab-FB1/GOMn_3O_4/ITO bioelectrode surface to remove the unbound ab-FB1 (pH 7.0). Non-specific sites on Ab-FB1/GOMn_3O_4/ITO bioelectrode were blocked with BSA (10 mL), and incubation was initiated at room temperature for 4 hours in a humidifying chamber. PBS with a pH of 7.0 was used to rinse the BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode, which were then kept at 4°C for further study. The procedure for preparing the BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode is described in Scheme 1.



Accepted Manuscript



Scheme 1. (a) Schematic of synthesis of GOMn_3O_4 nanocomposite and fabrication of immunoelectrode on the ITO electrode surface for FB1 detection, (b) illustrates about the EPD set-up used for the EPD of GOMn_3O_4 nanocomposite onto ITO electrode surface under the applied voltage for a period of time followed by surface activation, immobilization and detection of FB1 using the electrochemical set-up of three electrode system comprised of working, counter and reference electrode.

3. Results and Discussions

3.1 Structural analysis of GOMn_3O_4 nanocomposite

Figure 1(a) shows the XRD of synthesized GO sheets. The characteristic intense peak appeared at 10.67° corresponding to the (001) plane and data were recorded between the 5° to 80° range at a 2θ angle. The interlayer spacing of GO was calculated using Bragg's law Equation [1]:

$$n\lambda = 2d\,Sin\theta \qquad [1]$$

Where, n is the order of diffraction, λ is the wavelength of the X-ray, d is the interlayer spacing, and θ is the diffraction peak. Using the above equation, the interplanar distance between the graphene sheets was found to be 0.828 nm [34]. Figure 1(b) shows the XRD data of Mn_3O_4 nanoparticles and Figure 1(c) shows the XRD data of GOMn_3O_4 nanocomposites. It was observed that the XRD pattern of Mn_3O_4 consists of several diffraction peaks at 13.82°, 29.13°, 31.32°, 32.60°, 36.33°, 44.69°, 51.01°, 58.76°,

Accepted Manuscript

60.13°, and 64.88°, which were corresponding to the lattice planes of (101), (112), (103), (211), (004), (220), (105), (321), (224) and (400), respectively. Tetragonal-phase Mn_3O_4 nanoparticles matched well with the reported JCPDS number 80-0382 [35-38]. The average crystallite size (D) was calculated to be 20.92 nm using the Scherrer equation, described in Equation [2]:

$$D = \frac{K\lambda}{\beta COS(\theta)} \qquad [2]$$

Where, K= Crystallite shape-factor = 0.9, wavelength (λ) of X-Rays = 1.5406 Å, β = full width at half maximum (FWHM) of the diffraction peaks, and θ = Bragg's diffraction angle. The XRD peaks of Mn_3O_4 matched well with the XRD peaks of $GOMn_3O_4$ nanocomposite. All the same XRD peaks are appeared in Figure 1 (c) attributed to the presence of Mn_3O_4 nanoparticles in the GO sheets. However, the intense peak of GO disappeared in the XRD pattern of the nanocomposite. It indicated that the Mn_3O_4 was mainly formed on the surface of graphene [39-44].

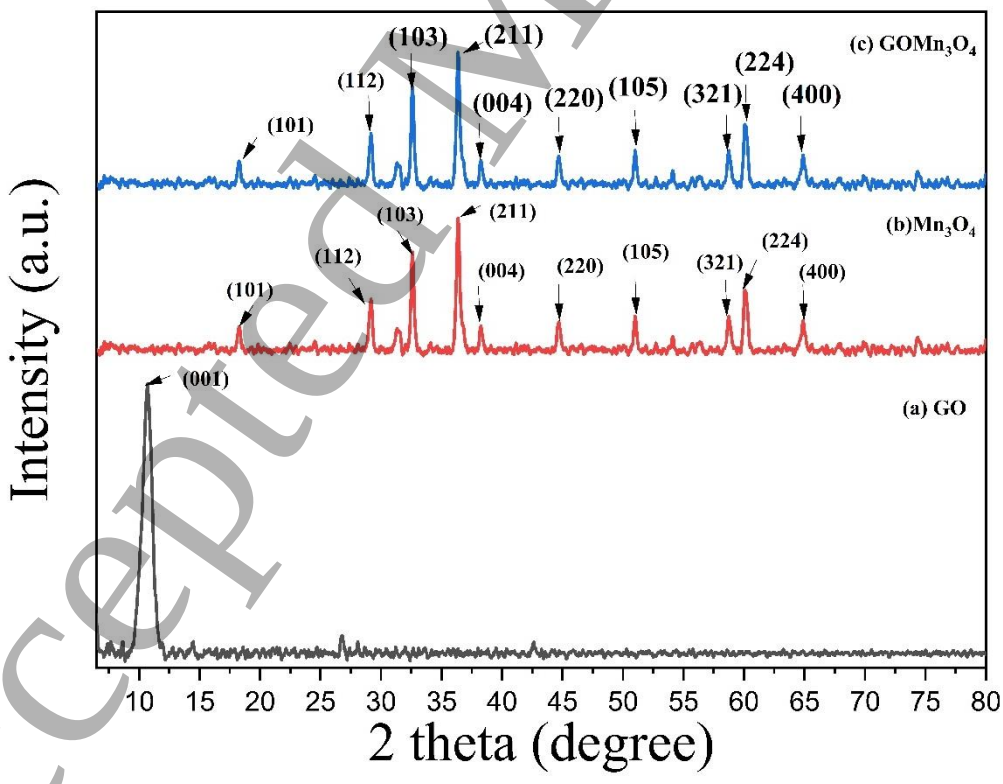


Figure 1: XRD spectrum of (a) GO, (b) Mn_3O_4, and (c) $GOMn_3O_4$ nanocomposite

Accepted Manuscript

The $GOMn_3O_4$ nanocomposite pure phase development was confirmed by the Raman spectrum depict in Figure 2. Figure 2(a) shows the synthesized GO contains two characteristics bands i.e. D and G-bands, corresponding to 1346 and 1598 cm^{-1}, respectively [45,46]. The intensity ratio of the D-band to the G-band is 0.84, which is less than 1, which shows a lower defect concentration. Raman spectrum of synthesized Mn_3O_4 nanoparticles was recorded and depicted in Figure 2(b), a very sharp peak was observed at 634 cm^{-1} which corresponds to the breathing vibration of divalent manganese ions in tetrahedral coordination [47]. Figure 2(c) illustrates the Raman spectrum of synthesized $GOMn_3O_4$ nanocomposite. The D-band and G-band shifted from 1346 to 1351, 1598 to 1591 cm^{-1} respectively, is due to the specific interaction of GO with Mn_3O_4. The sharp peak at 647 cm^{-1} corresponds to A_{1g} breathing vibration of the Mn-O tetrahedral spinel structure. At 305 and 360 cm^{-1}, there were also two additional oscillation peaks seen. The presence of the Mn-O bending mode was indicated by the first two low-intensity peaks, 305 and 360 cm^{-1} [48-49].

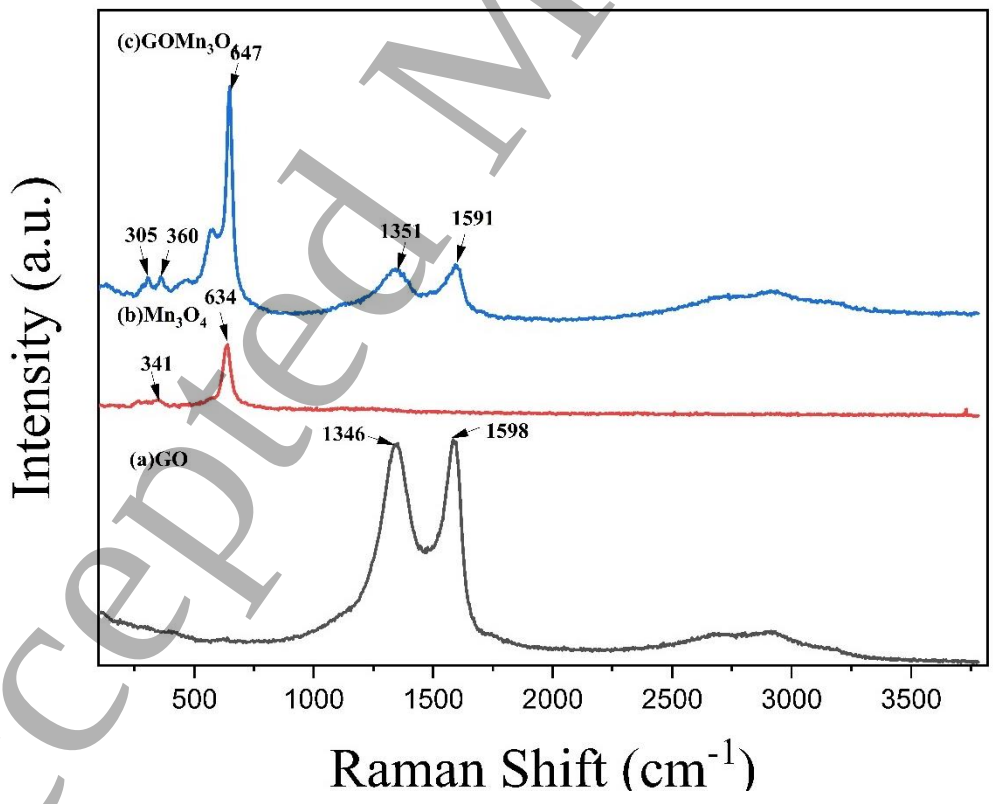


Figure 2: Raman spectrum of (a) GO, (b) Mn_3O_4, and (c) $GOMn_3O_4$ nanocomposite

Accepted Manuscript

Figure 3 shows the FE-SEM of EPD GO onto ITO-coated glass electrode surface. The appearance of GO sheets onto surface of ITO electrode confirms successful deposition of the materials by EPD technique. It can be seen that individual graphene sheets stack together and appear smooth over the ITO electrode surface. EPD technique is simple and fast for depositing the materials by applying certain voltage and time in comparison to the drop-cast method that have a lot of limitations including poor uniformity of the film, limited coverage area, leaching of nanomaterials. EPD is an electrochemical process which is used to migrate the colloidal suspended charged particles or molecules/ions towards their respective electrode surface either to cathode or anode by applying an appropriate voltage in a period of time into two-electrode cell. EPD technique has been used to attain a controlled thickness and to obtain a homogenous morphology of the film on electrode surface that play a significant role in obtaining the better electrochemical sensing performance. It is a simple, economical and eco-friendly technique that used for uniform deposition of materials onto the conducting surface. Therefore, for deposition of GO materials, EPD technique was used and deposited onto the ITO electrode surface under the optimized parameters of 60 volt for 5 minutes. Moreover, it can be seen that graphene sheets form flat like sheet structure and appeared smooth. A clear appearance of GO sheets over the ITO glass substrate confirms the synthesis and deposition of GO.

Accepted Manuscript

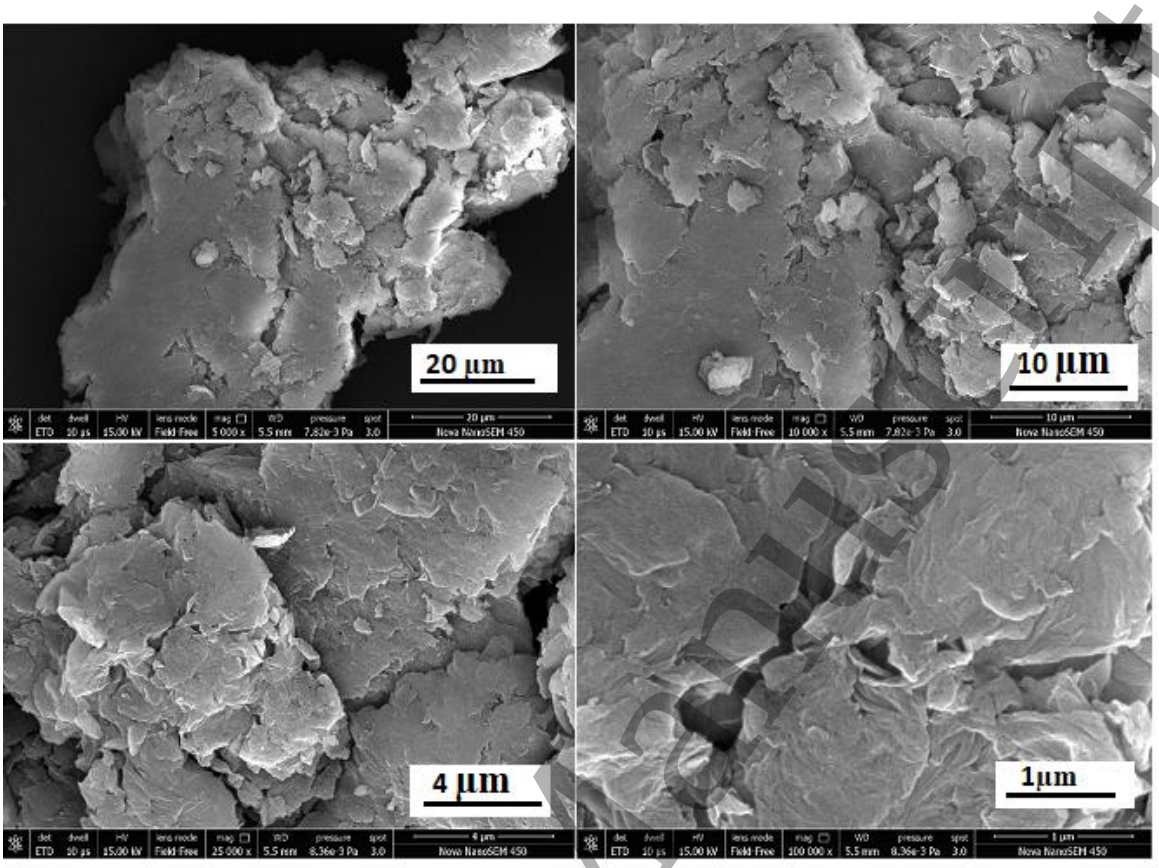


Figure 3: FE-SEM Images of synthesized GO

FTIR Study of GO, Mn_3O_4, $GOMn_3O_4$

The FTIR study of GO, Mn_3O_4, and $GOMn_3O_4$ was observed using potassium bromide (KBr) pellet in the 400 to 4000 cm^{-1} wavelength range, as shown in Figure 4. FTIR spectrum of GO sample shows in Figure 4(a) accorded well with the previous works [30]. There were various oxygen configurations in the structure. They include hydroxyl (namely phenol, C-OH) (3050-3800 cm^{-1}), ketonic species (C = O) (1600-1650 cm^{-1}), including C-OH vibration at 1083 cm^{-1}. In Figure 4(b) of Mn_3O_4, three main absorption peaks occurred. The peak at 3430 cm^{-1} corresponds to O-H stretching and bending modes of absorbed water [50]. The peak at 615 cm^{-1} was attributed to the vibration of manganese species (Mn^{3+}) in a tetrahedral site. The peak at 487 cm^{-1} was attributed to the vibration of manganese species (Mn^{3+}) in an octahedral site [51]. FTIR spectrum was recorded for the $GOMn_3O_4$ composite also and result shown in Figure 4(c). The peak appeared at 3447 cm^{-1} corresponds to O-H stretching and bending modes of

Accepted Manuscript

absorbed water and peaks appeared at 1590 and 1411 cm^{-1} attributed to the C=O and C=C bonding, respectively. The two sharp intense peaks occurred at 618 and 496 cm^{-1} corresponds to vibration of manganese species (Mn^{3+}) in a tetrahedral site and in an octahedral site, respectively.

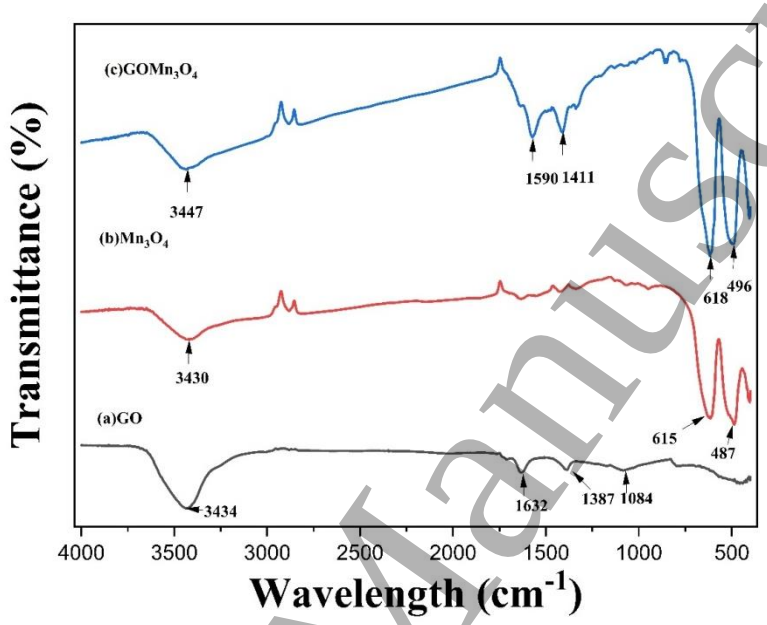


Figure 4: FTIR spectrum of (a) GO, (b) Mn_3O_4, and (c) $GOMn_3O_4$ nanocomposite

Elemental and morphological study of $GOMn_3O_4$ nanocomposite

Figure 5 illustrates the shape and crystallinity of the $GOMn_3O_4$ nanocomposite as determined by TEM (a-c). The distribution of Mn_3O_4 nanoparticles on the graphene sheet surface is depicted in Figure 5 (a, b). The arrow points to a layer of graphene oxide with Mn_3O_4 nanoparticles decorated onto the GO sheets and average size of the nanoparticle found to be 52.38±1.40 nm. Figure 5 (c) shows an HR-TEM image of the $GOMn_3O_4$ nanocomposite showing the thickness of the gap. The HR-TEM images were assigned to the (101) lattice plane of Mn_3O_4 tetragonal phase, and the d spacing was calculated to be 0.45 nm.

Figure 5(d) illustrates EDX imaging used to find the elemental composition of the $GOMn_3O_4$ nanocomposite. The carbon (C), manganese (Mn), and oxygen (O) elements could be found in the $GOMn_3O_4$ nanocomposite hybrid structure, according to EDX

Accepted Manuscript

spectra. The elements Mn and O were observed in the EDX spectra due to the presence of Mn_3O_4, while C was seen due to the presence of graphene nanosheets.

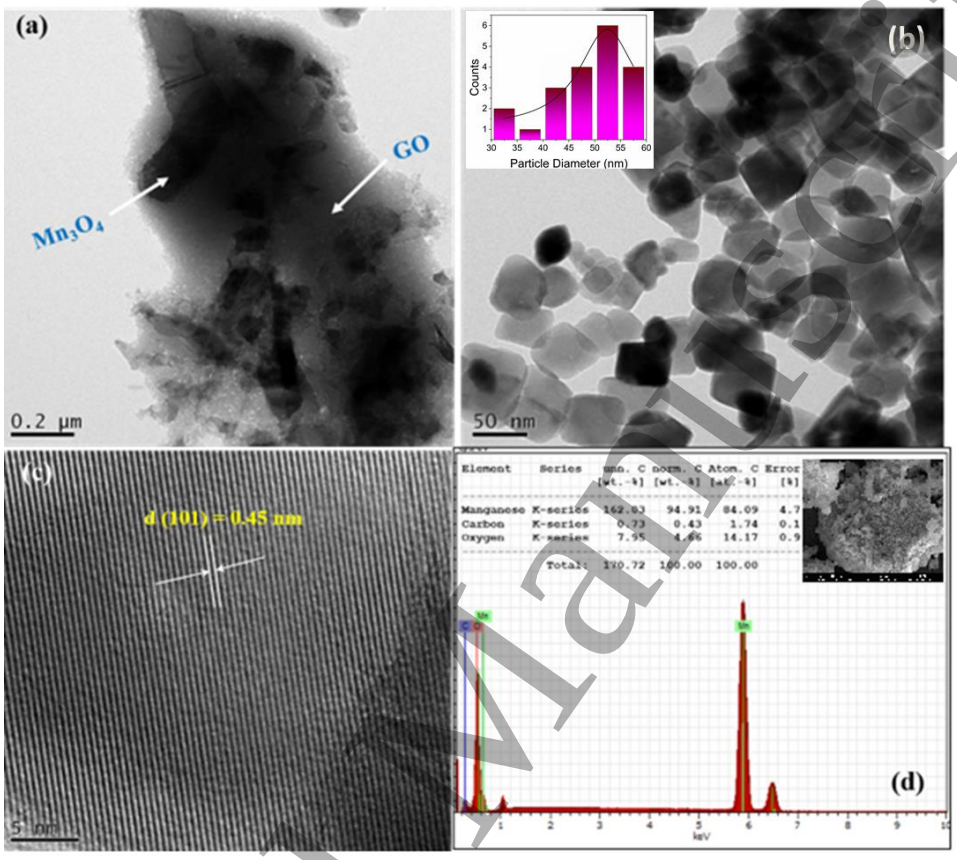


Figure 5: (a, b) describes the HR-TEM images of $GOMn_3O_4$ nanocomposite [inset image: particle size column graph], (c) HR-TEM lattice fringe image of $GOMn_3O_4$, and (d) EDX image of $GOMn_3O_4$ nanocomposite.

3.2 FTIR study of ab-FB1 onto $GOMn_3O_4$/ITO

The immobilization of ab-FB1 and BSA on the surface of the $GOMn_3O_4$/ITO immunoelectrode were investigated using attenuated total reflection (ATR) FTIR spectroscopy in the 400 to 4000 cm^{-1} range, results are shown in Figure 6. The FTIR spectra of the $GOMn_3O_4$/ITO immunoelectrode are shown in Figure 6(i) includes distinctive peaks at 530, 612, 1085, and 1412 cm^{-1}. The two peaks appeared at 1085 and 1412 cm^{-1} correspond to C-O stretching vibration and the -OH group in the carboxylic acid, respectively present in the graphene oxide, while the other two peaks at 530 and

Accepted Manuscript

612 cm^{-1} correspond to the presence of Mn-O bonds in the Mn_3O_4 nanoparticles. In Figure 6(ii), the peaks appeared at 1068, 1570, and 1642 cm^{-1} were attributed to C-N stretching, NH_2 group deformation (amide-II bond), and primary amide N-H of the fabricated ab-FB1/GOMn_3O_4/ITO immunoelectrodes [52]. These results were obtained due to the presence of ab-FB1 on the surface of the GOMn_3O_4/ITO immunoelectrode. The FTIR spectrum of the BSA/Ab-FB1/GOMn_3O_4/ITO immunoelectrode was also performed to confirm that the non-specific binding site of the antibody on the electrode surface was blocked by BSA [53]. As shown in Fig. 6(iii), the FTIR spectrum showed a characteristic decrease in the peak intensity of ab-FB1, and the peak position of the Mn-O bond was shifted from 530 to 536 cm^{-1} at Mn_3O_4 nanoparticle [54]. The above studies showed that ab-FB1 and BSA could be successfully immobilized on the surface of GOMn_3O_4/ITO electrode.

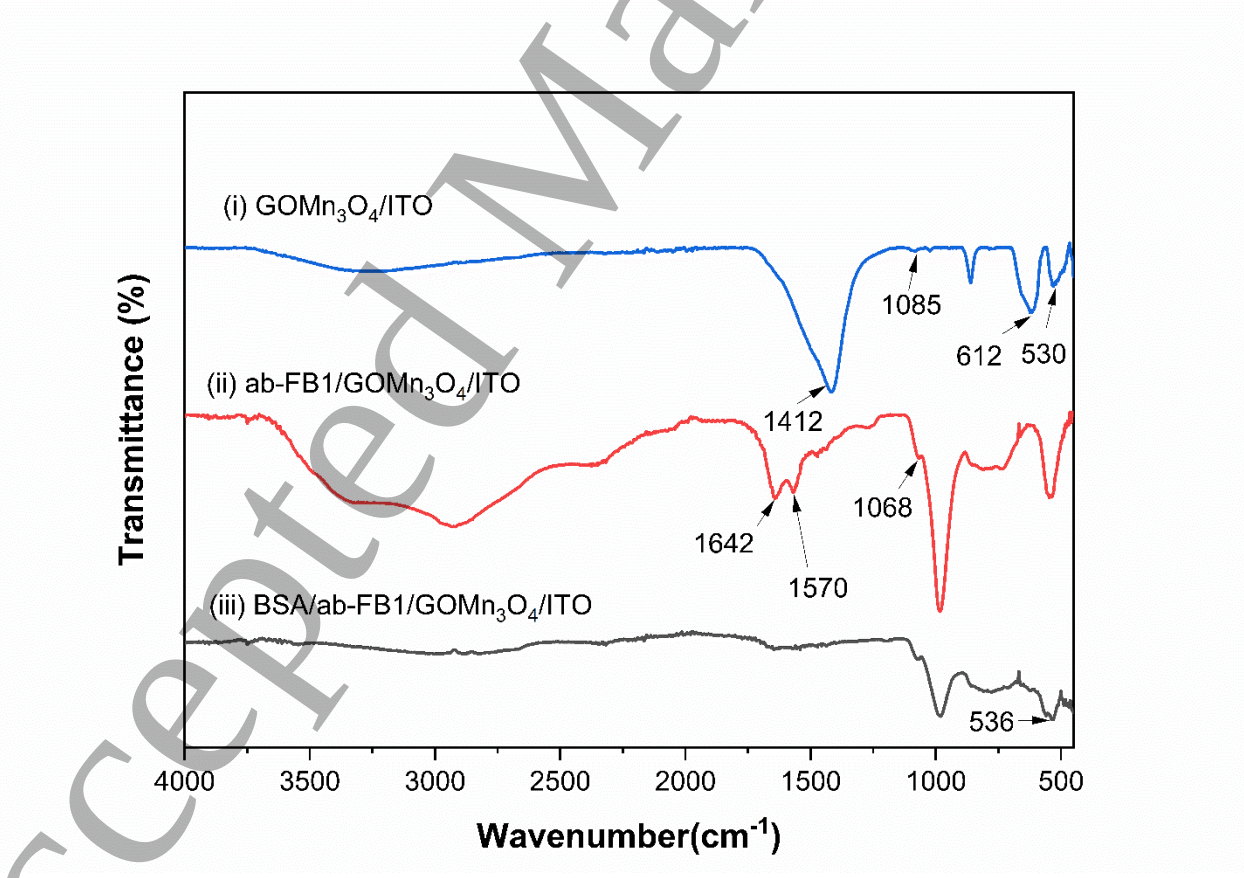


Figure 6: (a) FTIR spectrum of (i) GOMn_3O_4/ITO, (ii) abFB1/GOMn_3O_4/ITO, and (iii) BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode.

Accepted Manuscript

3.3 Electrochemical Study

3.3.1 Optimization conditions

To carry out the electrochemical biosensing studies of the fabricated immunosensor BSA/ab-FB1/GOMn_3O_4/ITO for FB1 detection, optimization studies are important to optimize the certain parameters for the fabrication of sensor. Therefore, few parameters such as pH study, incubation time study were carried out using the BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode.

Figure 7(a), presents the electrochemical responses of immunosensor BSA/ab-FB1/GOMn_3O_4/ITO at different pH ranging from 6.0 to 8.0 of PBS containing $[Fe(CN)_6]^{3-/4-}$ due to the effect of pH onto the protein molecules. As by changing from lower to higher pH value i.e. from acidic to basic medium of the electrolytic solution PBS containing $[Fe(CN)_6]^{3-/4-}$, can lead to alter the conformation of ab-FB1 and BSA present on the surface of immunoelectrode. Therefore, it is important to evaluate the specific pH for the fabrication of immunosensor. By adjusting the pH from 6.0 to 8.0 and DPV electrochemical responses were recorded under the potential range from -0.2 to +0.4 V and pH study results are depicted in Figure 7(a). The maximum current was obtained at pH 7.0 in the specified pH range (6.0 to 8.0). Since the composition of ab-FB1 might be altered by acidic or basic media, therefore, pH 7.0 of PBS containing $[Fe(CN)_6]^{3-/4-}$ was used for the fabrication of immunosensor for FB1 detection.

The incubation time was optimized at 5 minutes intervals of time to analyze the interaction between immunoelectrode (BSA/ab-FB1/GOMn_3O_4/ITO) and FB1. The incubation study plot is shown in Figure 7(b). The interaction of immunoelectrode (BSA/abFB1/GOMn_3O_4/ITO) and FB1 (10 pg mL^{-1} fixed concentration) mycotoxin in PBS containing $[Fe(CN)_6]^{3-/4-}$ buffer solution was studied using the DPV method. The decrease in peak current for immunoelectrode was observed and saturated after 35 minutes. As a result, the incubation period for the FB1 analyte to interact with the ab-FB1 immobilized on the surface of the GOMn_3O_4/ITO electrode was optimized to 35 minutes.

Accepted Manuscript

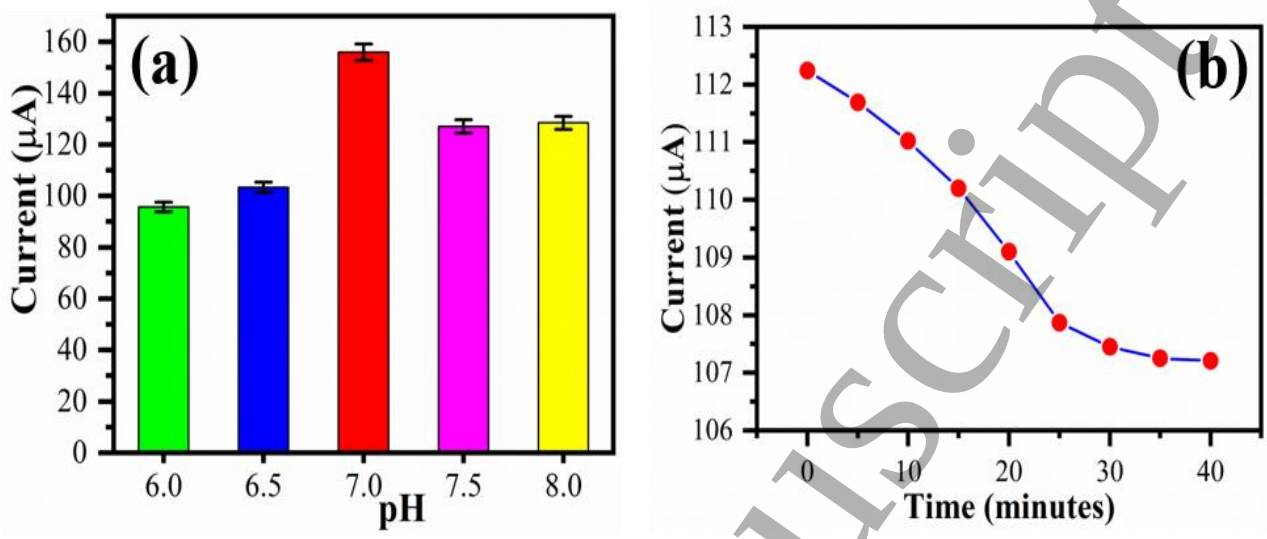


Figure 7: (a) The effect of different pH from 6.0 to 8.0 in PBS containing $[Fe(CN)_6]^{3-/4-}$, and (b) incubation plot for BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode, in PBS containing $[Fe(CN)_6]^{3-/4-}$.

3.3.2 An electro-kinetic scan rate study

The electro-kinetic interface parameters were measured for electrodes GOMn_3O_4/ITO, abFB1/GOMn_3O_4/ITO, and BSA/abFB1/GOMn_3O_4/ITO using cyclic voltammetry technique with changing the scan rate from 20 to 100 mV/s which is shown in the Figure 8 (a-c). Peak anode (I_a) and cathode (I_c) redox currents were calculated for the electrodes GOMn_3O_4/ITO, abFB1/GOMn_3O_4/ITO, and BSA/abFB1/GOMn_3O_4/ITO immunoelectrode at 50 mV/s.

The inset Figure (8a, b, c) shows a plot of the anode (I_a) and cathode (I_c) currents that are linearly proportional to the square root ($\upsilon^{1/2}$) of the scan rate. The peak current ratio (I_a/I_c) was determined to be 1.11, 1.12, 1.13, which corresponds to the immunoelectrodes GOMn_3O_4/ITO, ab-FB1/GOMn_3O_4/ITO, and BSA/ab-FB1/GOMn_3O_4/ITO, respectively. Thus, the obtained results of the current ratio indicate the quasi-reversible electron transfer process. Furthermore, the peak potential was also determined using the relation $\Delta E_p = E_{pa} - E_{pc}$ for the immunoelectrode and found to be 0.35 V. The determined potential is higher than ab-FB1/GOMn_3O_4/ITO electrode which ascribes the efficient electron transfer between immunoelectrode and redox solution.

Accepted Manuscript

The diffusivity (D) of the immunoelectrode was found using the Randles-Sevcik equation described below (equation [3]):

$$I_p = (2.69 \times 10^5) C\, n^{3/2} D^{1/2} \nu^{1/2} A \qquad [3]$$

Where, I_P is the peak current of the electrode, C is the concentration of the electrolyte, n is the number of electrons, A represents the surface area of the working electrode, and ν is the scan rate (mV/s). The diffusivity for BSA/abFB1/GOMn_3O_4/ITO immunoelectrode is found to be $2.08 \times 10^{-11} \mathrm{cm^2 s^{-1}}$ which indicates that at the electro-analytical interface, the rate of electron transfer is higher.

The availability of binding sites present at the immunoelectrode was calculated with the help of the above-mentioned diffusion-controlled parameter and the area (A_e) was determined using the Randles-Sevcik equation, described below in equation [4]:

$$Ae = \frac{S}{(2.69 \times 10^5) n^3 C D^{1/2}} \qquad [4]$$

Where, S is the slope of the straight line obtained from the linear plot of I_{pa} *vs* scan rate. The obtained results show a 4.11 mm^2 area which indicates the presence of more binding sites per unit volume at the immunoelectrode.

Further, the concentration at the surface of ionic species (I*) was also determined using Brown-Anson equation [5] which is as follows:

$$Ip = \frac{n^2 F^2 I^* A V}{4RT} \qquad [5]$$

Where, F is Faraday constant (96,485 C mol^{-1}), R stands for gas constant (8.314 J mol^{-1} K^{-1}), T stands for temperature (300 K), and the meaning of additional symbols is the same as that given in the previous equation. The determined result was found to be $2.94 \times 10^{-9} \mathrm{mol\,cm^{-2}}$ for the BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode which is higher for the reactive species at the interface.

Accepted Manuscript

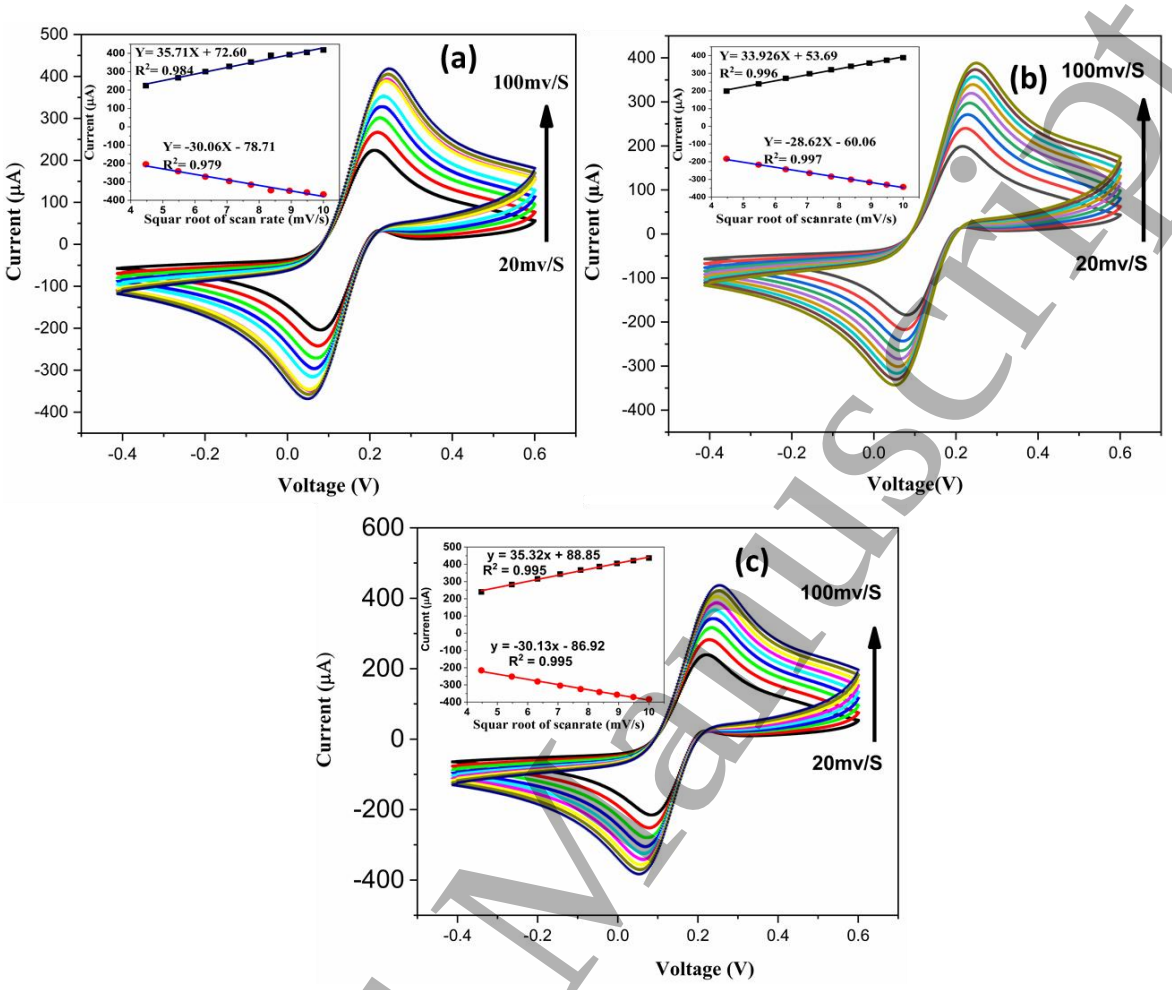


Figure 8: Scan rate study from 20 to 100 mV/s and linear plot between the square root of scan rate *vs* peak current of (a) $GOMn_3O_4$/ITO, (b) ab-FB1/$GOMn_3O_4$/ITO, and (c) BSA/ab-FB1/$GOMn_3O_4$/ITO immunoelectrode, respectively.

Electrochemical characterization studies of $GOMn_3O_4$/ITO, ab-FB1/$GOMn_3O_4$/ITO and BSA/ab-FB1/$GOMn_3O_4$/ITO surfaces were carried out using DPV technique. The DPV spectra of $GOMn_3O_4$/ITO, abFB1/$GOMn_3O_4$/ITO, and BSA/ab-FB_1/$GOMn_3O_4$/ ITO immunoelectrodes are shown in Figure 9 under the potential range from -0.2 to 0.4 V PBS containing $[Fe(CN)_6]^{3-/4-}$ at pH 7.0. The magnitude of peak current values of (i) $GOMn_3O_4$/ITO is higher than (ii) ab-FB_1/$GOMn_3O_4$/ITO. This happens due to the large surface-to-volume ratio, excellent adhesion, good electron conductivity and high electro-catalytic activity of the material. After the immobilization of ab-FB_1 over the $GOMn_3O_4$/ITO electrode surface, peak current value decreased is due to the presence

Accepted Manuscript

of heavier bulky molecule. Then, modified electrode surface was washed with BSA protein to remove the non-binding sites therefore, the peak current increased.

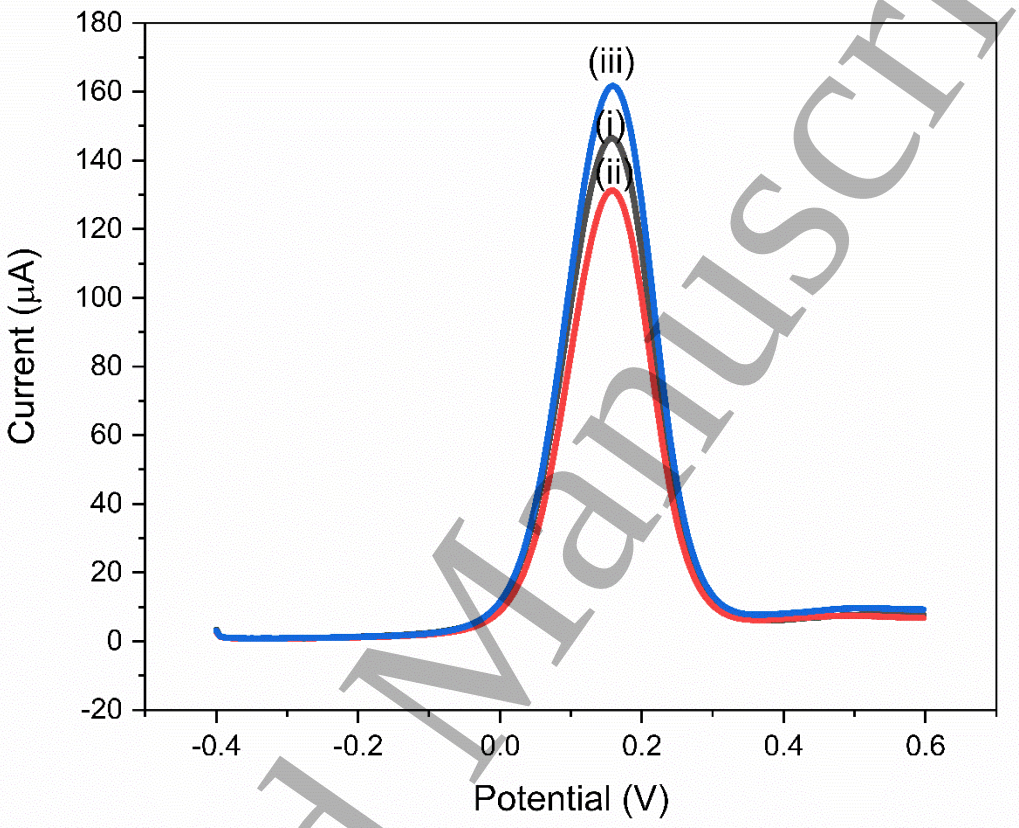


Figure 9: DPV response study of the (i) $GOMn_3O_4$/ITO, (ii) ab-FB1/$GOMn_3O_4$/ITO, (iii) BSA/ab-FB1/$GOMn_3O_4$/ITO electrode surface

Response study

Electrochemical biosensing response studies were carried out using DPV technique in a three electrode system containing BSA/ab-FB_1/$GOMn_3O_4$/ITO as a working electrode, reference electrode (Ag/AgCl) and counter electrode (Pt) in a PBS containing $[Fe(CN)_6]^{3-/4-}$ in the potential range from -0.2 to 0.4 V. The DPV technique is a voltammetry method that is used to make the electrochemical measurements response in which a pulsed potential (ΔE) is apply within a linear potential scanning using fixed increments. The electrochemical response is similar to the first derivate of a linear

Accepted Manuscript

voltammogram where a peak originated from the Faradaic reaction and can be identified near to $E_{1/2}$. The DPV measurements can be used to study the redox properties of extremely low concentration of analyte because effect of the charging capacitive current can be minimized, so high sensitivity is achieved than the linear scanning method and only faradaic current is extracted so the electrode reactions can be analysed more precisely.

To carried out the sensing response study, different concentration of analyte FB1 were analysed by DPV technique to acquire higher sensitivity, better limit of detection, better resolution for lower concentration of analyte. Therefore, electrochemical biosensing studies were carried out using the FB1 concentration range from 1 pg mL^{-1} to 800 ng mL^{-1} and result depicted in Figure 10. Figure 10(a) shows that the current response proportionally decreased as by increasing the FB1 concentration ranging from 1 pg mL^{-1} to 800 ng mL^{-1}. As by increasing the concentration of FB1 onto the immunoelectrode surface of BSA/ab-FB_1/$GOMn_3O_4$/ITO interaction taking place between the immobilized ab-FB1 and FB1 which leads to the formation of electrically insulating immunocomplex between ab-FB1 and FB1 that hinders the electron transfer between electrode and electrolyte correspondingly current response signal decreased. Figure 10 (b) shows the calibration curve between FB1 concentration and current response. Linear calibration plot presents the linear correlation followed between current responses with range of concentration of analyte from 1 pg mL^{-1} to 800 ng mL^{-1}. The fabricated immunoelectrode electrochemical response in terms of sensitivity was calculated using the following of Sensitivity = slope/area of the electrode, assuming a 0.25 cm^{-2} electrode area. The immunoelectrode sensitivity was determined to be 10.08 μA mL ng^{-1} cm^{-2}, with a regression coefficient (R^2) of 0.97. The measurements were carried out in triplicate under the same experimental conditions and signal-to-noise ratio (S/N) is three. LOD was calculated using the standard formula of 3σ/m, (σ is the standard deviation and m is the slope) and LOD value found to be 0.195 pg mL^{-1}. Better and remarkable increase in the biosensing response for FB1 detection is happens due to the use of $GOMn_3O_4$ nanocomposite in comparison to other surface matrixes as summarized in the Table 1.

Accepted Manuscript

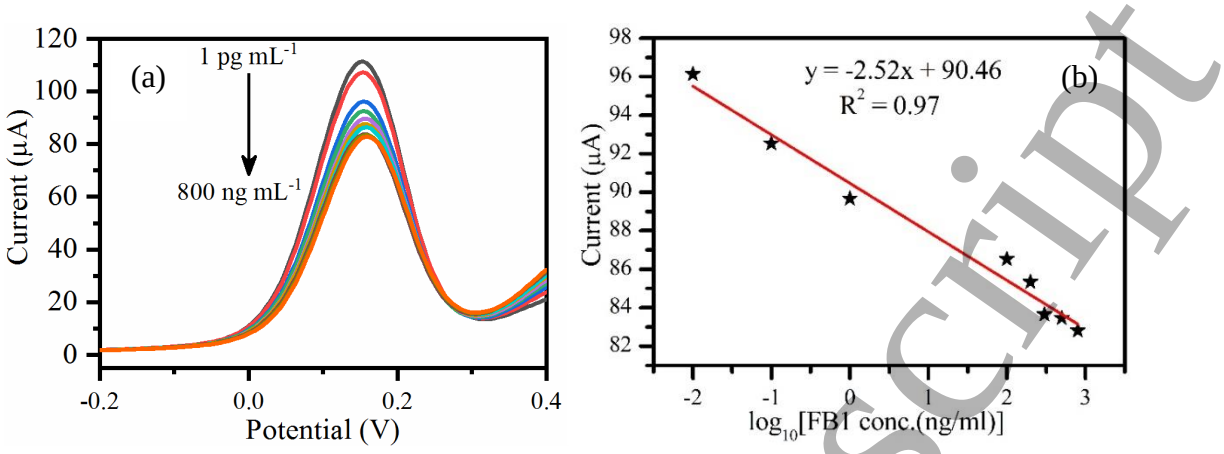


Figure 10: (a) DPV biosensing response, and (b) a linear plot of the BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode, respectively against FB1 in the concentration range from 1 pg mL^{-1} to 800 ng mL^{-1}.

The biosensing response obtained from the current fabricated immunoelectrode BSA/ab-FB1/GOMn_3O_4/ITO for FB1 detection is comparable to the previously reported articles for the FB1 detection and findings are summarized in Table 1. GOMn_3O_4 nanocomposite increased surface-to-volume ratio [55], improve loading of ab-FB_1 over GOMn_3O_4/ITO surface, provide good electrical conductivity [56], electron transfer efficacy [57,58], available abundant of functional groups for conjugation with antibodies thereby it enhance the sensitivity, linear detection range, limit of detection, specificity, selectivity and stability of the immunosensor. The electrochemical biosensing signal response parameters are found to be better in comparison to the previously published work due to the integration of GOMn_3O_4 nanocomposite for the fabrication of immunosensor for FB1 detection.

Table 1: illustrates the comparison of present work with other reported methods for the detection of FB_1

Method	**Procedure**	**Detection Range (ng mL^{-1})**	**Sensitivity (µA mL ng^{-1} cm^{-2})**	**LOD**	**Ref.**
HIS/MIP/Au/IO	CV	50nM-500 µM	-	1.07 nM	[19]
CeO_2-Ag/ abFB1/ITO	CV	0.01 -100	7.33	1.5 pg mL^{-1}	[20]
FB1/MIP	CV	0.001 -500	9.09	0.32 pg mL^{-1}	[3]

Accepted Manuscript

FB1/FnanoBIT assay	HPLC	0.533-6.81	-	0.079 ng mL^{-1}	[21]
FB1/GO-Au	CV	0.01 – 0.01	-	0.01 ng mL^{-1}	[22]
FB1/GO-CdS	CV	0.01 – 1000	-	4.7 pg mL^{-1}	[23]
AuNPs-MB-UCNPs	FRET	0.01 -100	-	0.01 ng mL^{-1}	[24]
FB1-HRP/ abFB1/ MB and protein G/ CSPE	DPV	0.73 -11.2	-	0.33 ng mL^{-1}	[25]
FELISA/CAT/QDs	ELISA	0.39-12.50	-	0.22 ng mL^{-1}	[26]
GCE/SWCNTs/CS	EIS	0.01-1000	-	2 pg mL^{-1}	[27]
GNB/LFIA	TOF-MS	3.9-125	-	1.76 ng mL^{-1}	[28]
MBs/DNA	Fluorescence	0.1-40	-	0.061 ng mL^{-1}	[29]
BSA/ab-FB1/GOMn_3O_4/ ITO	CV	0.001-800	10.08	0.195 pg mL^{-1}	This work

HIS: Histamine; IO: Inverse opal electrode; abFB1: Antibody FB1; MIP: Molecularly imprinted polymer; FnanoBIT assay: Ab-fumonisin antibody nano luminescent binary technology; GO-CdS: Graphene oxide– cadmium sulfide; ECL-ELISA: Enhanced chemiluminescent/Enzyme-linked immunosorbent assay; AuNPs-MB-UCNPs: Gold nanoparticles-methylene blue- upconversion nanoparticles; PPy/ErGO-SPE: Polypyrrole/electrochemically reduced graphene oxide-screen printed electrode; HRP: Horseradish peroxidase; Ab-FB1/MB: Abbody-FB1/methylene blue; protein G/CSPE: protein guanine/carbon screen-printed electrode; BSA: Bovine serum albumin; FELISA: Fluorescence enzyme-linked immunosorbent assay; CAT: catalase; QDs: Mercaptopropionic acid-modified CdTe quantum dots; GCE: Glass carbon electrode; SWCNTs: single-walled carbon nanotubes; CS: Chitosan; GNB: Gold nanobeads; LFIA: Lateral flow immunoassay; TOF-MS: Time-off-flight mass spectrometry; MBs: Magnetic beads

Accepted Manuscript

3.3.3 Repeatability, Reproducibility, control, and interferent study

Reproducibility studies of immunosensor BSA/ab-FB1/GOMn_3O_4/ITO were performed on five different immunoelectrodes using the DPV technique. Figure 11(a) displays current response obtained from five numbers of immunoelectrodes. The column graph demonstrated remarkable consistency for a specific immunoelectrode under the same settings, with a relative standard deviation (RSD) of 2.49%. Due to the stability and reproducibility of the immunoelectrodes, almost the same current response with minimal variation was obtained with the five different immunoelectrodes.

Figure 11(b) shows the repeatability study of the fabricated immunoelectrode BSA/ab-FB1/GOMn_3O_4/ITO which was measured at a FB1 concentration of 5 ng mL^{-1} using the DPV technique. A virtually steady current was obtained with a RSD of 1.06 % by five consecutive readings. This study verified the designed BSA/ab-FB1/GOMn_3O_4/ITO immunosensor have remarkable repeatability.

It is important to investigate the interference study of the fabricated immunosensor to measure the specificity and selectivity of the sensor towards the FB1 only. Therefore, interference study was carried out with various interference analytes to see the influence of other analyte onto the electrode surface of BSA/ab-FB1/GOMn_3O_4/ITO. For this study, various interferents including other mycotoxins such as AFB1, OchA, and DON, and other carbohydrates like cellulose, fructose, and starch which are found in the food sample were used and DPV response obtained as shown in Figure 11(c). The selectivity study was measured with the addition of 20 µL of interferents in the electrolytic solution in presence of FB1 mycotoxins and their response was recorded using the DPV technique. The obtained results showed the fabricated immunosensor is highly selective towards the presence of FB1. These results indicate that ab-FB1 are highly specific for the FB1.

Also, specificity studies were carried out to investigate the binding of FB1 only with ab-FB1 [Figure 11(d)]. To perform this study, control study using the GOMn_3O_4/ITO electrode tested with the various series of concentration of FB1 from 1 pg mL^{-1} to 800 ng mL^{-1}. It is observed that the no change in current signal found indicates that the FB1

Accepted Manuscript

analyte is highly specific for the immobilized ab-FB1 onto the GOMn_3O_4/ITO electrode.

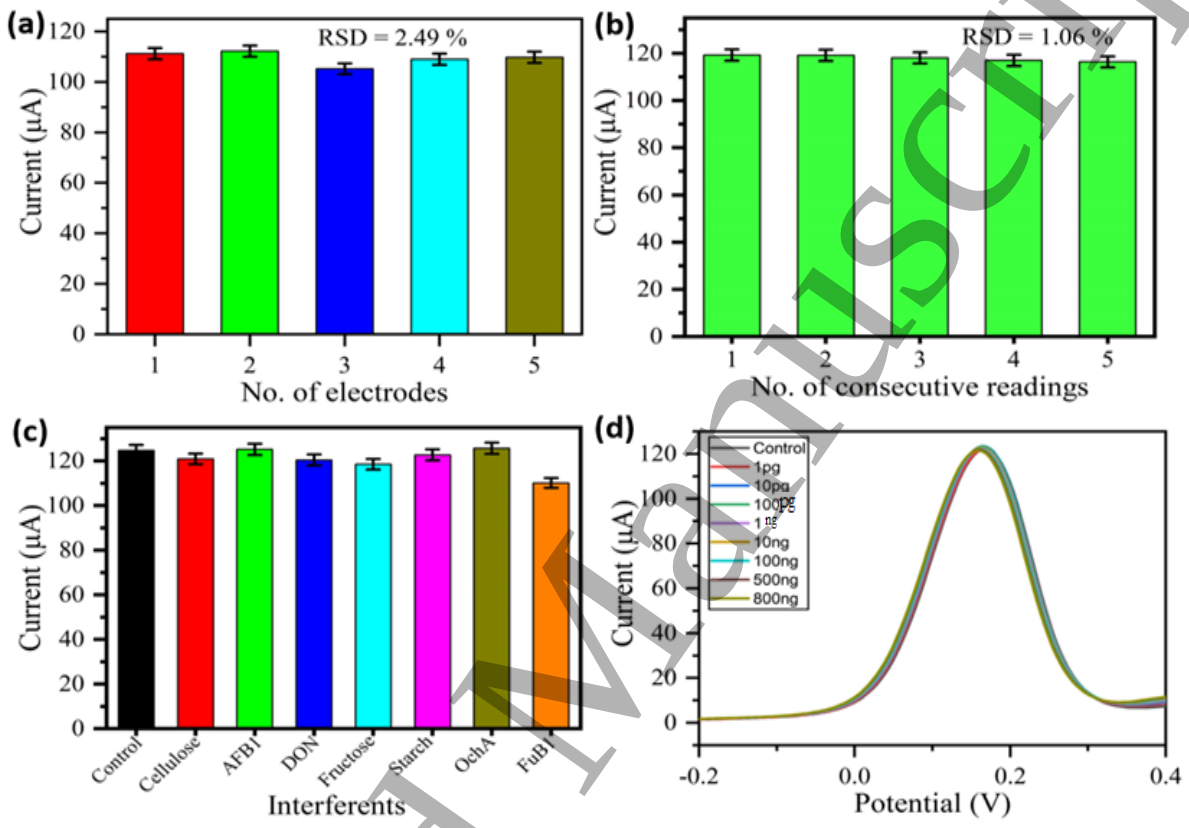


Figure 11: (a) Reproducibility plot for five different immunoelectrodes, (b) shows the repeatability study measured for five different consecutive readings with 1.06 % RSD value, (c) describes the interferent study plot with various interfering analytes, and (d) depicts the GOMn_3O_4/ITO control measurement plot with different concentrations range from 1 pg mL^{-1} to 800 ng mL^{-1}.

3.3.4 Spiked Study

To determine the real sample study, BSA/ab-FB_1/GOMn_3O_4/ITO immunoelectrode responses were observed in spiked maize samples with different FB1 concentrations. The fresh sweet maize was purchase from the local market and crushed thoroughly and diluted in the D.I. water. Further, the diluted maize extract was taken and mixed with different concentration of FB1 (including 1 pg mL^{-1}, 10 pg mL^{-1}, 100 pg mL^{-1}, and 1 ng

Accepted Manuscript

mL^{-1}, 100 ng mL^{-1}, 300 ng mL^{-1} and 500 ng mL^{-1}) and measurement was performed using DPV technique under similar conditions. Figure 12 shows the linear calibration graph is plotted for both spiked and sensing response obtained for various FB1 concentration from 1 pg mL^{-1}to 500 ng mL^{-1} in sweet maize samples. Linear regression coefficient (R^2) is a measure that provides information about goodness of fit in the linear plot. Regression is a statistical measure of how well the regression line approximates the actual data. The R^2 values were calculated for the spiked and sensing results and found to be 0.82 for spiked and 0.85 for sensing studies. The lower R^2 (0.82) for spiked sample studies is due to the small change in the current values with changing the concentration of FB1 in spiked maize samples while in case of FB_1 sensing studies, current is decreasing linearly with the increase of FB1 concentration resulting improved in R^2 (0.85). However, it is observed that in both spiked sample and with FB1, current values is decreased as by increasing the concentration of FB1 antigen indicates that that fabricated immunoelectrode BSA/ab-FB1/GOMn_3O_4/ITO shows specific binding for FB_1 analyte in sensing and spiked samples. The percentage of recovery rate was calculated and found to be in the range of 98.91% to 123.3 %. These results show that an excellent recovery and this fabricated immunoelectrode BSA/ab-FB1/GOMn_3O_4/ITO can be used for the FB1 detection.

Accepted Manuscript

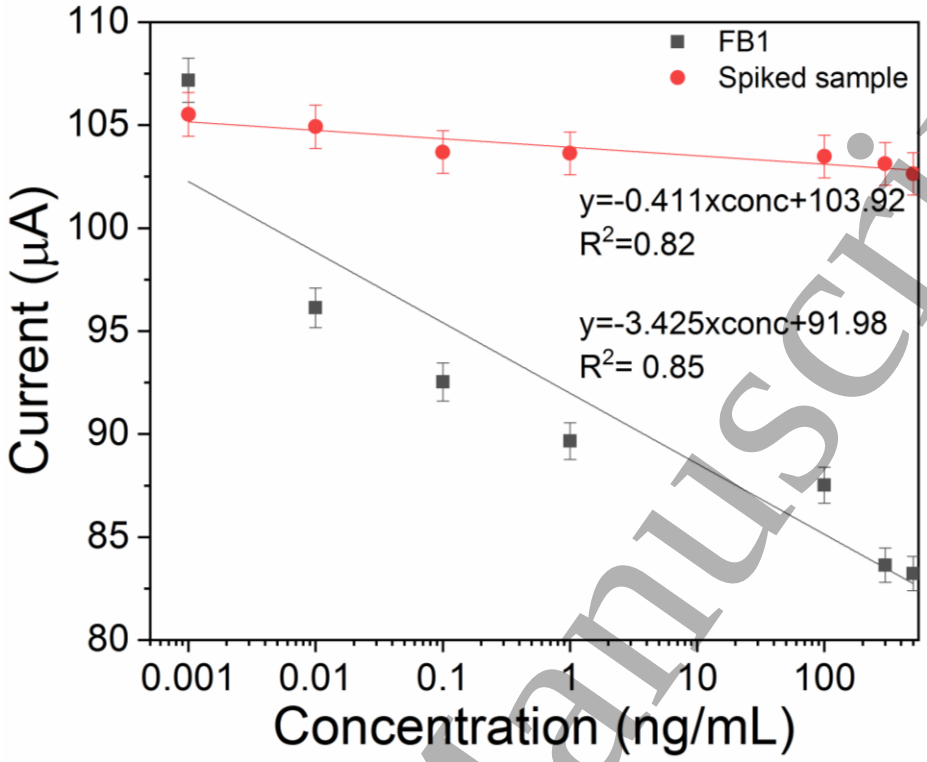


Figure 12: Linear plot of FB1 and spiked sample studies for concentrations of 1 pg mL^{-1}, 10 pg mL^{-1}, 100 pg mL^{-1}, and 1 ng mL^{-1}, 100 ng mL^{-1}, 300 ng mL^{-1} and 500 ng mL^{-1} performed in the sweet maize samples using the fabricated BSA/ab-FB1/$GOMn_3O_4$/ITO immunoelectrode

4. Conclusions

BSA/ab-FB1/$GOMn_3O_4$/ITO immunoelectrode was fabricated for the sensitive and precise detection of FB1 in sweet corn samples in the concentration range from 1 pg mL^{-1} to 800 ng mL^{-1}. $GOMn_3O_4$ nanocomposite was successfully synthesized by wet chemical co-precipitation method. XRD and HR-TEM characterization techniques were used and validated its tetragonal phase of structure. A $GOMn_3O_4$ nanocomposite thin film was developed on an ITO-coated glass surface using an EPD method. Ab-FB1 and BSA were immobilized on $GOMn_3O_4$/ITO electrodes and the immobilization was confirmed using DPV and FTIR spectroscopy. The fabricated immunoelectrode exhibit a high sensitivity of 10.08 µA mL $ng^{-1}cm^{-2}$ and LOD was found to be 0.195 pg mL^{-1} with R^2 of 0.97 for FB1 detection. The FB1 spiked maize samples recovery rates were found in the range from 98.91% to 123.3%. This shows that the BSA/ab-

Accepted Manuscript

FB1/GOMn_3O_4/ITO immunoelectrode system is fully applicable to detect FB1 in actual samples. Various naturally occurring interferences in food (along with FB1), such as AFB1, OchA, DON, cellulose, fructose, and starch, were used in interference studies to confirm the specific detection of FB1. With five electrodes and continuous measurements, the BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode were likewise shown to be very reliable and reproducible. In a controlled study, the GOMn_3O_4/ITO electrode showed no tendency to detect FB1, and the BSA/ab-FB1/GOMn_3O_4/ITO immunoelectrode confirmed the specific interaction with FB1.

Declaration of Competing Interest

There is no conflict to declare.

Acknowledgment

The authors thank the Advanced Instrumentation Research Facility (AIRF), JNU for assisting with the XRD, SEM-EDAX, Raman and TEM characterization. The authors are thankful to MNIT, Jaipur FE-SEM image analysis. The authors are very grateful to the Lab D, Special Center for Nanosciences, JNU for FTIR, and electrochemical data analysis. We thank to Indian Council of Medical Research, New Delhi, India and Department Health Research, Ministry of Health and Family Welfare, Government of India (Grant No: R.11013/46/2021-GIA/HR) and Department of Science and Technology, New Delhi, India for financial support.

References:

[1] Oswald, I.P., Desautels, C., Laffitte, J., Fournout, S., Peres, S.Y., Odin, M., Le Bars, P., Le Bars, J. and Fairbrother, J.M., 2003. Mycotoxin fumonisin B1 increases intestinal colonization by pathogenic Escherichia coli in pigs. *Applied and Environmental Microbiology*, 69(10), pp.5870-5874.

[2] Bennett, J.W., 2003. Klich. M. Mycotoxins. *Clinical Microbiology Reviews*, 16(3), pp.497-516.

Accepted Manuscript

[3] Singh, A.K., Lakshmi, G.B.V.S., Fernandes, M., Sarkar, T., Gulati, P., Singh, R.P. and Solanki, P.R., 2021. A simple detection platform based on molecularly imprinted polymer for AFB1 and FuB1 mycotoxins. *Microchemical Journal*, *171*, p.106730.

[4] Rheeder, J.P., Marasas, W.F., Thiel, P.G., Sydenham, E.W., Shephard, G.S. and Van Schalkwyk, D.J., 1992. Fusarium moniliforme and fumonisins in maize in relation to human esophageal cancer in Transkei Vol 82, No.3.

[5] Plattner, R.D., Norred, W.P., Bacon, C.W., Voss, K.A., Peterson, R., Shackelford, D.D. and Weisleder, D., 1990. A method of detection of fumonisins in maize samples associated with field cases of equine leukoencephalomalacia. *Mycologia*, *82*(6), pp.698-702.

[6] Wang, S., Quan, Y., Lee, N. and Kennedy, I.R., 2006. Rapid determination of fumonisin B1 in food samples by enzyme-linked immunosorbent assay and colloidal gold immunoassay. *Journal of Agricultural and food chemistry*, *54*(7), pp.2491-2495.

[7] Shephard, G.S., Thiel, P.G. and Sydenham, E.W., 1992. Determination of fumonisin B1 in plasma and urine by high-performance liquid chromatography. *Journal of Chromatography B: Biomedical Sciences and Applications*, *574*(2), pp.299-304.

[8] Hartl, M. and Humpf, H.U., 1999. Simultaneous determination of fumonisin B1 and hydrolyzed fumonisin B1 in maize products by liquid chromatography/electrospray ionization mass spectrometry. *Journal of agricultural and food chemistry*, *47*(12), pp.5078-5083.

[9] Peña-Bahamonde, J., Nguyen, H.N., Fanourakis, S.K. and Rodrigues, D.F., 2018. Recent advances in graphene-based biosensor technology with applications in life sciences. *Journal of Nanobiotechnology*, *16*, pp.1-17.

[10] Li, D., Zhang, W., Yu, X., Wang, Z., Su, Z. and Wei, G., 2016. When biomolecules meet graphene: from molecular level interactions to material design and applications. *Nanoscale*, *8*(47), pp.19491-19509.

[11] Gagrani, A., Ding, B., Wang, Y. and Tsuzuki, T., 2019. pH dependent catalytic redox properties of Mn_3O_4 nanoparticles. *Materials Chemistry and Physics*, *231*, pp.41-47.

Accepted Manuscript

[12] Hosny, N.M. and Dahshan, A., 2012. Facile synthesis and optical band gap calculation of Mn_3O_4 nanoparticles. *Materials Chemistry and Physics*, *137*(2), pp.637-643.

[13] Şahin, B., Aydin, R. and Cetin, H., 2020. Tuning the morphological, structural, optical and dielectric properties of hausmannite (Mn_3O_4) films by doping heavy metal lead. *Superlattices and Microstructures*, *143*, p.106546.

[14] Gao, J., Lowe, M.A. and Abruna, H.D., 2011. Spongelike nanosized Mn_3O_4 as a high-capacity anode material for rechargeable lithium batteries. *Chemistry of Materials*, *23*(13), pp.3223-3227.

[15] Reza, K.K., Singh, N., Yadav, S.K., Singh, M.K. and Biradar, A.M., 2014. Pearl-shaped highly sensitive Mn_3O_4 nanocomposite interface for biosensor applications. *Biosensors and Bioelectronics*, *62*, pp.47-51.

[16] Gund, G.S., Dubal, D.P., Patil, B.H., Shinde, S.S. and Lokhande, C.D., 2013. Enhanced activity of chemically synthesized hybrid graphene oxide/Mn3O4 composite for high performance supercapacitors. *Electrochimica Acta*, *92*, pp.205-215.

[17] Khan, H.A., Lee, Y.K., Shaik, M.R., Alrashood, S.T. and Ekhzaimy, A.A., 2022. Nanocomposites of Nitrogen-Doped Graphene Oxide and Manganese Oxide for Photodynamic Therapy and Magnetic Resonance Imaging. *International Journal of Molecular Sciences*, 23(23), p.15087.

[18] Chen, J., Bai, Z., Li, X., Wang, Q., Du, J., Lu, R. and Liu, X., 2022. In-situ synthesis of reduced graphene oxide wrapped Mn_3O_4 nanocomposite as anode materials for high-performance lithium-ion batteries. *Ceramics International*, 48(21), pp.31923-31930.

[19] Ahmed, S., Ansari, A., Haidyrah, A.S., Chaudhary, A.A., Imran, M. and Khan, A., 2022. Hierarchical Molecularly Imprinted Inverse Opal-Based Platforms for Highly Selective and Sensitive Determination of Histamine. *ACS Applied Polymer Materials*, *4*(4), pp.2783-2793.

[20] Dhiman, T.K., Lakshmi, G.B.V.S., Dave, K., Roychoudhury, A., Dalal, N., Jha, S.K., Kumar, A., Han, K.H. and Solanki, P.R., 2021. Rapid and label-free electrochemical detection of fumonisin-B1 using a microfluidic biosensing platform

Accepted Manuscript

based on Ag-CeO_2 nanocomposite. *Journal of The Electrochemical Society*, *168*(7), p.077510.

[21] Alsulami, T., Nath, N., Flemming, R., Wang, H., Zhou, W. and Yu, J.H., 2021. Development of a novel homogeneous immunoassay using the engineered luminescent enzyme NanoLuc for the quantification of the mycotoxin fumonisin B1. *Biosensors and Bioelectronics*, *177*, p.112939.

[22] Zheng, Y., Shi, Z., Wu, W., He, C. and Zhang, H., 2021. Label-free DNA electrochemical aptasensor for fumonisin B 1 detection in maize based on graphene and gold nanocomposite. *Journal of Analytical Chemistry*, *76*, pp.252-257.

[23] Mao, L., Ji, K., Yao, L., Xue, X., Wen, W., Zhang, X. and Wang, S., 2019. Molecularly imprinted photoelectrochemical sensor for fumonisin B1 based on GO-CdS heterojunction. *Biosensors and Bioelectronics*, *127*, pp.57-63.

[24] Wu, S., Duan, N., Li, X., Tan, G., Ma, X., Xia, Y., Wang, Z. and Wang, H., 2013. Homogenous detection of fumonisin B1 with a molecular beacon based on fluorescence resonance energy transfer between NaYF4: Yb, Ho upconversion nanoparticles, and gold nanoparticles. *Talanta*, *116*, pp.611-618.

[25] Jodra, A., López, M.Á. and Escarpa, A., 2015. Disposable and reliable electrochemical magnetoimmunosensor for Fumonisins simplified determination in maize-based foodstuffs. *Biosensors and Bioelectronics*, *64*, pp.633-638.

[26] Lu, T., Zhan, S., Zhou, Y., Chen, X., Huang, X., Leng, Y., Xiong, Y. and Xu, Y., 2018. Fluorescence ELISA based on CAT-regulated fluorescence quenching of CdTe QDs for sensitive detection of FB 1. *Analytical Methods*, *10*(48), pp.5797-5802.

[27] Yang, X., Zhou, X., Zhang, X., Qing, Y., Luo, M., Liu, X., Li, C., Li, Y., Xia, H. and Qiu, J., 2015. A highly sensitive electrochemical immunosensor for fumonisin B1 detection in maize using single-walled carbon nanotubes/ chitosan. *Electroanalysis*, 27(11), pp.2679-2687.

[28] Chen, X., Miao, X., Ma, T., Leng, Y., Hao, L., Duan, H., Yuan, J., Li, Y., Huang, X. and Xiong, Y., 2021. Gold nanobeads with enhanced absorbance for improved sensitivity in competitive lateral flow immunoassays. *Foods*, *10*(7), p.1488.

[29] Qu, C., Zhao, L., He, X., Yu, S. and Wei, M., 2021. Magnetic beads–assisted fluorescence aptasensing approach based on dual DNA tweezers for detection of

Accepted Manuscript

ochratoxin A and fumonisin B 1 in wine and maize. *Analytical and Bioanalytical Chemistry*, *413*, pp.6677-6685.

[30] Zaaba, N.I., Foo, K.L., Hashim, U., Tan, S.J., Liu, W.W. and Voon, C.H., 2017. Synthesis of graphene oxide using modified hummers method: solvent influence. *Procedia Engineering*, *184*, pp.469-477.

[31] Singh, A.K., Dhiman, T.K., Lakshmi, G.B.V.S., Raj, R., Jha, S.K. and Solanki, P.R., 2022. Rapid and label-free detection of Aflatoxin-B1 via microfluidic electrochemical biosensor based on manganese (III) oxide (Mn_3O_4) synthesized by co-precipitation route at room temperature. *Nanotechnology*, *33*(28), p.285501.

[32] Nam, I., Kim, N.D., Kim, G.P., Park, J. and Yi, J., 2013. One step preparation of Mn_3O_4/graphene composites for use as an anode in Li ion batteries. *Journal of Power Sources*, *244*, pp.56-62.

[33] Ata, M.S., Liu, Y. and Zhitomirsky, I., 2014. A review of new methods of surface chemical modification, dispersion and electrophoretic deposition of metal oxide particles. *Rsc Advances*, *4*(43), pp.22716-22732.

[34] Ain, Q.T., Haq, S.H., Alshammari, A., Al-Mutlaq, M.A. and Anjum, M.N., 2019. The systemic effect of PEG-nGO-induced oxidative stress in vivo in a rodent model. *Beilstein Journal of Nanotechnology*, *10*(1), pp.901-911.

[35] Kanwal, A., Sajjad, S., Leghari, S.A.K. and Khan, M.N., 2021. Strong interfacial charge transfer between hausmannite manganese oxide and alumina for efficient photocatalysis. *Chinese Journal of Chemical Engineering*, 33, pp.147-159.

[36] Rani, B.J., Ravina, M., Ravi, G., Ravichandran, S., Ganesh, V. and Yuvakkumar, R., 2018. Synthesis and characterization of hausmannite (Mn_3O_4) nanostructures. *Surfaces and Interfaces*, 11, pp.28-36.

[37] Zawrah, M.F., El Fadaly, E.A., Khattab, R.M., Aly, M.H. and El Shafei, H., 2020. Synthesis and characterization of nano Mn_3O_4 and $LiMn_2O_4$ spinel from manganese ore and pure materials. *Ceramics International*, 46(11), pp.17514-17522.

[38] Bose, V.C., Maniammal, K., Madhu, G., Veenas, C.L., Raj, A.A. and Biju, V., 2015, February. DC electrical conductivity of nanocrystalline Mn_3O_4 synthesized through a novel sol-gel route. In IOP conference series: materials science and engineering (Vol. 73, No. 1, p. 012084). IOP Publishing.

Accepted Manuscript

[39] Nam, I., Kim, N.D., Kim, G.P., Park, J. and Yi, J., 2013. One step preparation of Mn_3O_4/graphene composites for use as an anode in Li ion batteries. *Journal of power sources*, *244*, pp.56-62.

[40] Zhu, L., Zhang, S., Cui, Y., Song, H. and Chen, X., 2013. One step synthesis and capacitive performance of graphene nanosheets/Mn_3O_4 composite. *Electrochimica Acta*, *89*, pp.18-23.

[41] Zhang, X., Sun, X., Chen, Y., Zhang, D. and Ma, Y., 2012. One-step solvothermal synthesis of graphene/Mn_3O_4 nanocomposites and their electrochemical properties for supercapacitors. *Materials Letters*, 68, pp.336-339.

[42] Li, Y., Qu, J., Gao, F., Lv, S., Shi, L., He, C. and Sun, J., 2015. In situ fabrication of Mn_3O_4 decorated graphene oxide as a synergistic catalyst for degradation of methylene blue. *Applied Catalysis B: Environmental*, 162, pp.268-274.

[43] Jiangying, Q., Feng, G., Quan, Z., Zhiyu, W., Han, H., Beibei, L., Wubo, W., Xuzhen, W. and Jieshan, Q., 2013. Highly atom-economic synthesis of graphene/Mn_3O_4 hybrid composites for electrochemical supercapacitors. *Nanoscale*, 5(7), pp.2999-3005.

[44] Wang, L., Li, Y., Han, Z., Chen, L., Qian, B., Jiang, X., Pinto, J. and Yang, G., 2013. Composite structure and properties of Mn_3O_4/graphene oxide and Mn_3O_4 /graphene. *Journal of Materials Chemistry A*, 1(29), pp.8385-8397.

[45] Khan, A., Habib, M.R., Jingkun, C., Xu, M., Yang, D. and Yu, X., 2021. New insight into the metal-catalyst-free direct chemical vapor deposition growth of graphene on silicon substrates. *The Journal of Physical Chemistry C*, *125*(3), pp.1774-1783.

[46] Khan, A., Cong, J., Kumar, R.R., Ahmed, S., Yang, D. and Yu, X., 2022. Chemical Vapor Deposition of Graphene on Self-Limited SiC Interfacial Layers Formed on Silicon Substrates for Heterojunction Devices. *ACS Applied Nano Materials*, *5*(12), pp.17544-17555.

[47] Zahn, D.T., 1999. Vibrational spectroscopy of bulk and supported manganese oxides. *Physical Chemistry Chemical Physics*, *1*(1), pp.185-190.

[48] Deljoo, B., Tan, H., Suib, S.L. and Aindow, M., 2020. Thermally activated structural transformations in manganese oxide nanoparticles under air and argon atmospheres. *Journal of Materials Science*, *55*(17), pp.7247-7258.

Accepted Manuscript

[49] Shah, H.U., Wang, F., Toufiq, A.M., Ali, S., Khan, Z.U.H., Li, Y., Hu, J. and He, K., 2018. Electrochemical properties of controlled size Mn_3O_4 nanoparticles for supercapacitor applications. *Journal of Nanoscience and Nanotechnology*, *18*(1), pp.719-724.

[50] Krishnan, B. and Mahalingam, S., 2017. Synthesis and characterization of Mn_3O_4 /BC nanocomposite and its antimicrobial activity. *Journal of Inorganic and Organometallic Polymers and Materials*, *27*, pp.275-284.

[51] Wang, Y., Hou, C., Lin, X., Jiang, H., Zhang, C. and Liu, G., 2021. Dye degradation studies of hausmannite manganese oxide (Mn_3O_4) nanoparticles synthesized by chemical method. *Applied Physics A*, *127*, pp.1-7.

[52] Bhardwaj, H., Pandey, M.K. and Sumana, G., 2019. Electrochemical Aflatoxin B1 immunosensor based on the use of graphene quantum dots and gold nanoparticles. *Microchimica Acta*, *186*, pp.1-12.

[53] Bhardwaj, H., Marquette, C.A., Dutta, P. and Sumana, G., 2020. Integrated graphene quantum dot decorated functionalized nanosheet biosensor for mycotoxin detection. *Analytical and Bioanalytical Chemistry*, *412*, pp.7029-7041.

[54] Karimi, M. and Eshraghi, M.J., 2017. One-pot and green synthesis of Mn_3O_4 nanoparticles using an all-in-one system (solvent, reactant and template) based on ethaline deep eutectic solvent. *Journal of Alloys and Compounds*, *696*, pp.171-176.

[55] Wang, L., Li, Y., Han, Z., Chen, L., Qian, B., Jiang, X., Pinto, J. and Yang, G., 2013. Composite structure and properties of Mn_3O_4/graphene oxide and Mn_3O_4 /graphene. *Journal of Materials Chemistry A*, *1*(29), pp.8385-8397.

[56] Gund, G.S., Dubal, D.P., Patil, B.H., Shinde, S.S. and Lokhande, C.D., 2013. Enhanced activity of chemically synthesized hybrid graphene oxide/Mn_3O_4 composite for high performance supercapacitors. *Electrochimica Acta*, *92*, pp.205-215.

[57] Jiangying, Q., Feng, G., Quan, Z., Zhiyu, W., Han, H., Beibei, L., Wubo, W., Xuzhen, W. and Jieshan, Q., 2013. Highly atom-economic synthesis of graphene/Mn_3O_4 hybrid composites for electrochemical supercapacitors. *Nanoscale*, *5*(7), pp.2999-3005.

[58] Wang, Y., 2015. Coprecipitated 3D nanostructured graphene oxide–Mn_3O_4 hybrid as anode of lithium-ion batteries. *Journal of Materials Research*, *30*(4), pp.484-492.

Accepted Manuscript

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Hindi Hit Songs Prediction Using Machine Learning Algorithms

Rohit Beniwal
Department of CSE
Delhi Technological University
Delhi, India
dr.rohitbeniwal@gmail.com

Satvik Batra
Department of CSE
Delhi Technological University
Delhi, India
satvikbatra@gmail.com

Shivank Gupta
Department of CSE
Delhi Technological University
Delhi, India
guptashivank54@gmail.com

Shashi Shekhar
Department of CSE
Delhi Technological University
Delhi, India
shashis2510@gmail.com

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205626

***Abstract*— The music industry has grown from independent composers releasing their self-published work to an entire extensive industry wherein an enormous amount of money is at stake depending on whether or not a particular song will be a hit on the charts. As a result, the music production industry has become more digitized and data-centric in order to try and find a mechanism for predicting a hit song. Extensive research work has been done to predict the salability of a particular song being put out by composers. However, the current research has been limited to Western English language songs only. Due to the lack of datasets and preliminary research on regional language songs, hit prediction for regional language songs has not been given any attention in the existing state of the art. Therefore, in this research paper, we predict the profitability and salability of a Hindi language song by predicting whether or not the song will be a hit or in other words whether or not the song will be listed on top music charts. High-level features of a song have been incorporated into our models as parameters that will affect the profitability of the song in question. To demonstrate this, we created a new dataset of Hindi songs. Moreover, we compared the results of various algorithms such as Random Forrest, Gaussian Naive Bayes, Multi-layer Perceptron (MLP), AdaBoost, and Logistic Regression optimized using SGD. Our results show that MLP achieved the best outcomes with a precision of 0.9429, recall of 0.8875, and F1 score of 0.8849.**

Keywords*— *AdaBoost, Gaussian Naive Bayes, Hindi Songs, Hit Song Prediction, Machine Learning, Random Forrest, Regional Language, MLP, SGD Classifier.

I. INTRODUCTION

The music industry is expanding at a rapid pace and it has become even more imperative for music production houses to be able to predict that the product that they are putting out in the industry will be able to make sustained profits. Like every other company or organization, minimization of losses and maximizing profits is one of the foremost concerns of music production houses. As a result, they turned towards data-driven analysis and Machine Learning technology to predict the salability of a particular song.

Any song released has various features associated with it when we talk strictly in technical terms. Some of the high-level features acting as a parameter for judging a particular song's success are danceability, loudness, acousticness, etc. In this research work, we took into account such features and how they are affecting a song's performance on the music charts.

Since music is an abstract form of art, it has become a fascinating topic of research in how to take into account the above-mentioned features and predict certain aspects associated with it. These features have also been used in past works to predict a song's performance on the charts and even what type of genre a particular song will be classified into. Although, many other factors also play a role in whether a song will be a hit or not ranging from social factors to economic factors. Many times, historical music preferences also play a role in what type of music or song will be a hit in what region of the world. In this research, we limited ourselves only to the audio features of the song.

Almost all existing work in the hit song prediction is done for the West and more specifically on English language songs. Due to the lack of focus on regional language songs, we decided to work on the prediction of a Hindi song's performance on the charts. For this, we manually created a new dataset following which, many Machine Learning algorithms are applied to this dataset. The performance of the different models is then compared based on various performance metrics. Thus the goal of our research is to successfully predict whether or not the song will be a hit by using various Machine Learning algorithms with the best accuracy and precision.

To achieve the above-mentioned goal of the research, we used Machine Learning algorithms like Random Forest, SGD classifier, AdaBoost, Gaussian Naive Bayes, and Multi-layer Perceptron (MLP). We then evaluated the results using different metrics like F1-score, Recall, and Precision .

The rest of the paper is organized as follows: Section 2 gives a summary of the existing research work on music's success prediction. Section 3 describes the recommended technique and methodology; Section 4 summarizes the result and analysis; Lastly, Section 5 concludes the research paper along with offering future directions.

II. RELATED WORK

Many attributes can be used to forecast hit songs, i.e., audio features and metadata aren't always the only factors affecting a song's performance on the charts. In contrast to audio, lyrical, and meta-data modalities.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:58 UTC from IEEE Xplore. Restrictions apply.

Fan et al. [1] in 2013 worked on predicting whether a song will be a hit or not, analyzing Chinese and British songs. They then compared them and how different factors influence different regional songs on whether the song will be a hit or not. It is also important to keep in mind that the regional Chinese songs were in the English language but produced and distributed in China. They ran trials utilizing pop music charts from China and the UK for 40 weeks. To determine whether a new song will turn out to be a hit, a support vector machine model and a time-weighted linear regression model are used using a set of 10 common auditory attributes. The study identified the most effective features for predicting hit songs in the Chinese and UK pop charts, with a focus on the top-performing characteristics. Compared to the UK experiment, the Chinese prediction model demonstrated greater precision in its results. They came to the conclusion that Chinese hit songs had audio feature characteristics that were very distinct from the hit songs of the UK.

Due to the dearth in presence of datasets during early iterations of research in hit song prediction, Gutiérrez et al. [2] in 2020 created their own new alternative dataset called SpotGenTrack Popularity Dataset (SPD). A multimodal full-circle, The HitMusicNet deep learning architecture was then introduced for forecasting the popularity of musical recordings. Due to the inclusion of three major modalities in the analysis, the proposed design fared better than earlier studies. The three formats are audio, lyrics, and meta-data.

Raza et al. [3] in 2020 worked on finding the most appropriate set of features that can be used in Machine Learning algorithms in order to predict a hit song, or whether a song will be a hit or not. They included an array of parameters employed on English song charts that could help predict a song being hit. The dataset was formed by aggregating data from various platforms, including technical attributes of songs and sentiment analysis of their lyrics. To determine if there is a secret method for predicting hit songs, sombre Machine Learning logs were employed. The Machine Learning algorithms employed were Random Forests, Naive Bayes, Decision Trees, and Logistic Regression. According to the findings, there are factors other than the technical data points that might foretell if a song would be successful. The authors came to the conclusion that predicting successful songs is not yet a data science endeavor.

Aum et al. [4] in 2022 collaborated on a project aimed at forecasting whether or not the songs will feature on the Billboard Hot 100 chart in a way predicting success of songs by analyzing the interaction between artists and their fans. This involved utilizing a dataset that included comments, posts, and quote tweets from the artist-fan community, as well as assessing the emotional sentiment and interaction patterns present in each piece of content. They achieved the best classification results using the XGBoost model fed with quote-tweet data. They offered a condensed method for monitoring and comprehending public impressions of the music fraternity through discussions on social media. Ultimately, they proposed that incorporating a quote tweet feature into artist-fan interactions on Twitter could be a viable tactic for enhancing the performance of songs and increasing their success.

The study of related works led us to the conclusion that a decent amount of research has been done generally in the field of hit song prediction. However, the majority of the research though has been restricted to the globally popular "pop" English/ Western music and songs. Regional language songs including Hindi have been neglected completely. Further research is required to arrive at an accurate and precise prediction of songs in the Hindi language specifically and regional languages around the globe generically and also make them more robust in the said domains. Therefore, in this research we focused on Hindi language songs and tried to predict whether or not it will be a hit song using various Machine Learning algorithms, to finally conclude which algorithm is giving the best results.

III. METHODOLOGY

The methodology for predicting the Hindi language hit songs using Machine Learning algorithms is built by evaluating it on a completely new custom-made dataset made specifically for this research with the help of different standard performance metrics. The following Fig. 1 shows the research methodology.

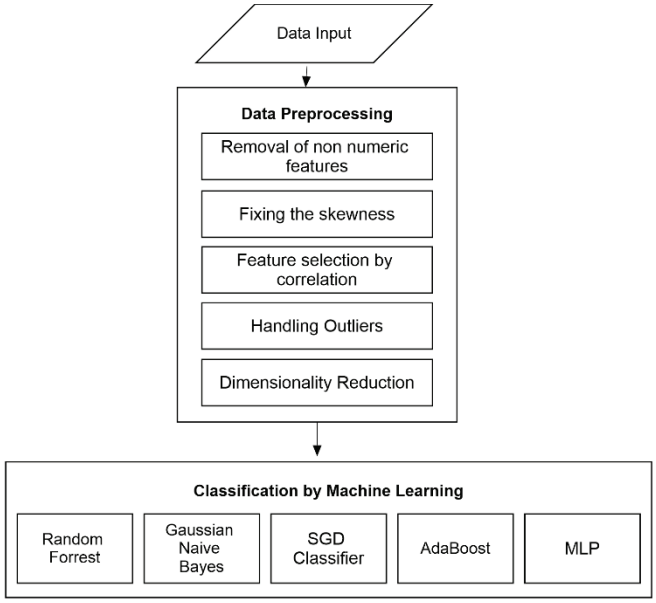


Fig. 1. The research methodology

The subsequent subsections explain the proposed research methodology used for predicting the Hindi language hit songs using Machine Learning algorithms.

A. Dataset

To implement the proposed research objective, we created our own dataset. The first part of the dataset was created by extracting the names of hit Hindi songs from the Billboard chart for top Hindi songs. To do this, we queried the Billboard API using billboard.py. Further we manually added the names of hit songs from various other music charts as no such API existed to extract them. This resulted in 834 data points of hit songs. We then picked the public dataset of Bollywood Songs available at kaggle.com [5]. With this, we further obtained 2337 data points. Common songs between Bollywood songs data and Billboard Data were removed. Finally, we were left with the data of about 2573 songs out of which 1964 Non-Hits and 609 were Billboard Hits. High-level audio features for each song were extracted using Spotipy [6] to query Spotify API. It included 23 high-level human understandable features such as danceability, loudness, energy, instrumentals, acoustic-ness, etc. Billboard

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:58 UTC from IEEE Xplore. Restrictions apply.

songs were classified as hits and labeled as 1 while the non-hits were labeled as 0 and the final dataset was created.

B. Implementation Details of Preprocessing Phase

Data preprocessing is carried out to eliminate the noise from the dataset. Here, data preprocessing includes multiple steps. From the 23 total features that we extracted, we removed non-numeric features and were left with 16 numeric features that are suitable for building classification models. We then verified the presence of any missing values, such as NaN or NULL, but found none as they had been eliminated during the data extraction process.

1) Standardization of Data

We observed the skewness of the features. Some features exhibited significant skewness from the normal distribution which is considered to be ideal, so the Yeo-Johnson power transform was applied to normalize and standardize the data. After transformation, the distributions of features were approximated to normal distributions with zero mean and unit standard deviation.

2) Feature Selection by Observing Correlation

We plot the correlation heatmap [7] of the features and make the following observations. Upon analyzing the heatmap, we made two key observations. Firstly, energy exhibited a strong positive relation with loudness (0.71) but a significant negative relation with acousticness (-0.66). Secondly, artist popularity and followers displayed a notable positive relation (0.48). To validate these correlations, we plotted pair-plots. We observed that energy and artist popularity increase nearly linearly with loudness and followers, respectively. The following fig. 2 shows the heatmap.

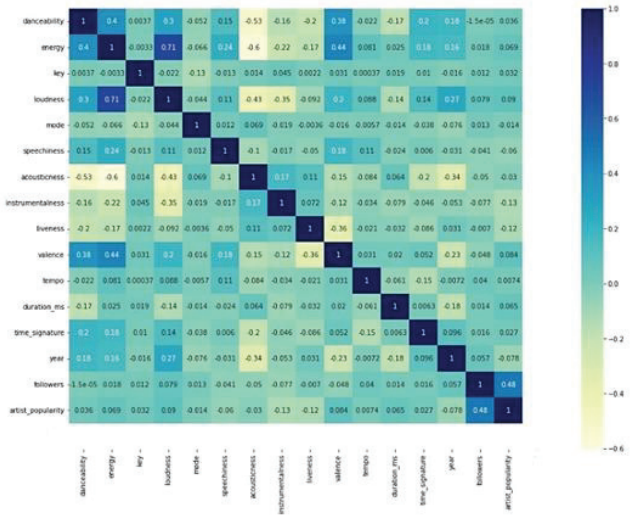


Fig. 2. Correlation Heatmap

As a result, we decided to eliminate energy and followers from our model since their effect on the output label can be effectively captured by loudness and artist popularity, respectively. We chose to remove energy due to its high correlation with two other variables. Furthermore, the distribution of loudness is closer to a normal distribution as compared to energy.

3) Handling Outliers

As the data has been standardized, we identified outliers for each feature based on their absolute Z-scores [8]. Specifically, we identified any data point that deviated from the mean by more than 2.6 standard deviations as an outlier. To determine the optimal threshold value for Z-score, we experimented with a range of values from 2 to 4 in increments of 0.1 (step size). Based on our results, we found that a Z-score threshold of 2.6 yielded the top outcomes.

After performing all the necessary preprocessing steps, our dataset was reduced to 14 features and 2280 data points.

4) Dimensionality Reduction

The technique of t-SNE [9] was employed to transform the data points into a two-dimensional representation, enabling us to visualize higher-dimensional data and comprehend the similarities among data points. "The pairwise similarity of the input objects and the corresponding low-dimensional points in the embedding are measured by two different distributions, and the t-Distributed stochastic neighbor embedding (t-SNE) minimizes the divergence between the two.". The following fig. 3 shows the t-SNE plot.

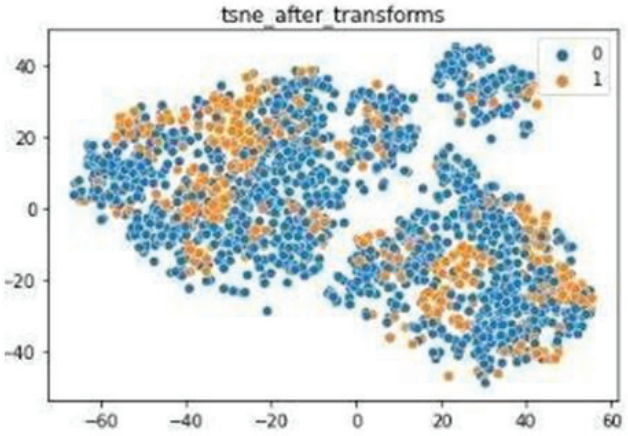


Fig. 3. t-SNE plot

C. Proposed Models and Details

The data was divided into two sets - a training set and a testing set - using a 70:30 ratio. The training set has 1596 samples and the testing set has 684 samples. We trained five well-known Machine Learning algorithms on our training dataset namely Random Forrest, Gaussian Naive Bayes, SGD Classifier, AdaBoost, and MLP.

Random Forrest is a type of ensemble model that aggregates outputs from multiple decision trees to generate a prediction. By combining the results of many decision trees, Random Forest can offer more robust and accurate predictions than a single decision tree.

Gaussian Naive Bayes is a probabilistic model that assumes feature independence to assign probabilities to output classes.

SGD Classifier is used to perform logistic regression using stochastic gradient descent and L1 regularization.

AdaBoost is a statistical classification algorithm that boosts the results of weak models.

MLP is a class of feedforward artificial neural networks. Strictly speaking, it comprises multiple layers of perceptrons with threshold activation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:58 UTC from IEEE Xplore. Restrictions apply.

IV. RESULT AND ANALYSIS

The methodology suggested in this study is implemented using the Python programming language. In addition, the following criteria were used to evaluate the classification results. Key Performance Indicators: precision, recall [10], F1 score & Accuracy. The degree of precision indicates how many samples that were expected to be positive turned out to be such. The ratio of true positive (TP) values to the total of all expected positive values is known as precision. (TP and FP).

Precision =True Positive / (True Positive + False Positive)

Recall quantifies the proportion of positive samples that are correctly predicted from the positive samples. Thus, it is explained as the proportion of true positive (TP) values to the total of all actual positive values. (TP and FN) .

Recall = True Positive / (True Positive + False Negative)

The harmonic mean of recall and precision is known as the F1-Score.

F1 score = 2 × Precision × Recall / (Precision + Recall)

A. *Result Details for Classification Phase*

We employed a 5-fold cross-validation fold [11] which was the best 'k' value in the range [2,10] to train our models and performed a grid search on the hyperparameters. We selected precision as the scoring metric during the grid search to keep down the number of false positives. The best model obtained from the grid search was then evaluated for precision, accuracy, F1 score, and recall. Table 1 below depicts the performance of the models and Fig 4 shows the comparative analysis.

TABLE I. PERFORMANCE EVALUATION OF CLASSIFICATION MODELS

Model	Precision	Accuracy	Recall	F1
RF	0.9122	0.8516	0.7810	0.8148
GNB	0.6622	0.3315	0.3688	0.3483
LR(SGD)	0.8732	0.8101	0.8055	0.7971
ADA	0.9225	0.8295	0.8639	0.8463
MLP	0.9429	0.8823	0.8875	0.8849

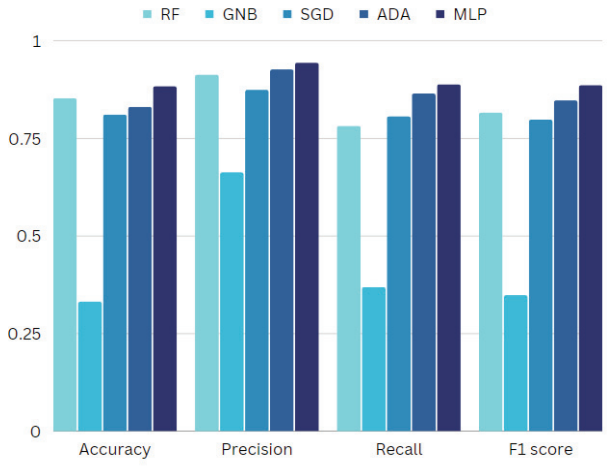


Fig. 4. Comparative analysis of results

B. *Analysis of Results*

MLP outperformed all other selected models across all performance metrics, which was expected given that Artificial Neural Networks (ANNs) are powerful function approximators.

As is evident from the t-SNE plot (Fig 3), the data is not linearly separable. Random Forrest worked well with such data and thus gave high precision and accuracy.

Gaussian Naive Bayes (GNB) gave the worst results. This is expected as Gaussian Naive Bayes assumes the features to be independent which is not true (apparent from the correlation heatmap Fig. 2)

Logistic regression (implemented using SGD) exhibited lower precision compared to Random Forrest, which is expected given that logistic regression works on the assumption that the data can be separated linearly. This assumption does not hold true in our scenario.

Relative feature importance was also observed in the Random Forest classifier. Fig. 5 below depicts the same.

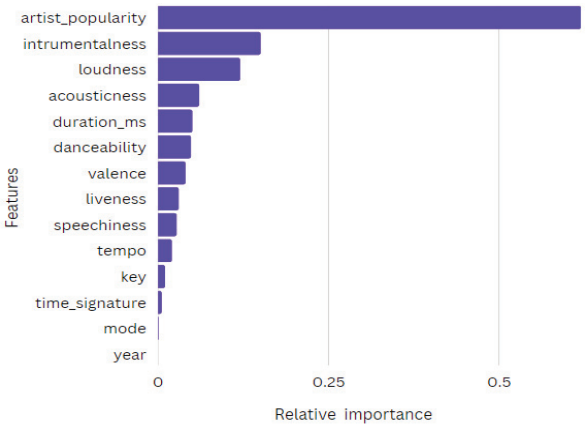


Fig. 5. Random Forest: Relative feature importance

The analysis of the feature importance clearly showed that the popularity of the artist had the highest relative importance by a significant margin, as expected. This finding is justifiable, as we anticipated that artist popularity would be largely determined by their past performance on similar metrics, such as being featured on Billboard charts or winning awards. Conversely, the year feature exhibited low importance, as hit songs were evenly dispersed across different years, and there was no discernible linear correlation between year and other features. This observation was corroborated by reviewing the correlation heatmap of the features with the labels.

V. CONCLUSION AND FUTURE SCOPE

In this research paper, we provided a Hindi hit song prediction using Machine Learning algorithms such as MLP, Random Forrest (RF), Adaboost, SGD Classifier, and Gaussian Naive Bayes (GNB). To demonstrate this, we created our own database using Billboard API and a list of songs at kaggle.com as mentioned in the Data Extraction process. Moreover, we compared the results of the MLP, RF, GNB, SGD, and ADA. The MLP model gave the best results and yielded a precision of 0.9429, a recall of 0.8875, and an F1 score of 0.8849.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:58 UTC from IEEE Xplore. Restrictions apply.

To enhance the accuracy of the model in future research, there are various paths that we can take.

In this project, we were able to implement high-level features like danceability, loudness, and vibrancy. In the future, to improve accuracy, we could incorporate low-level analysis wherein we can take a few seconds of sample audio clips of every song and extract features from that. Further, as we believe the lyrics of the songs are also a valuable factor in determining whether the song is going to be a hit or not, we can perform Natural Language Processing (NLP) on the text lyrics of the songs. There might be some complications in using NLP on Hindi lyrics but that is the challenge one would need to overcome. We could incorporate the project into a comprehensive software product and application so that it could be used by music labels and artists in predicting the chances of a new song becoming a hit.

In conclusion, our research has demonstrated the effectiveness of Machine Learning algorithms in identifying hit songs in the Hindi Language and has highlighted the need for further research in other low-resource languages such as Bengali, Chines, Malay, Thai, Vietnamese, etc. The developed models have the potential to be used in practical applications to detect and identify Hit Songs in the Hindi language.

REFERENCES

[1] J. Fan and M. Casey, "[PDF] study of Chinese and UK hit songs prediction: Semantic scholar," [PDF] Study of Chinese and UK Hit Songs Prediction Semantic Scholar, 01-Jan-1970.[Online].Available: https://www.semanticscholar.org/paper/Study-of-Chinese-and-UK-Hit-Songs-Prediction-Fan-Casey/3d4037c2817391df6c2b628bb8cef6ca39e57341. [Accessed: 21-Apr-2023].

[2] D. Martin-Gutierrez, G. Hernandez Penaloza, A. Belmonte-Hernandez, and F. Alvarez Garcia, "A multimodal end-to-end deep learning architecture for music popularity prediction," *IEEE Access*, vol. 8, pp. 39361–39374, 2020.

[3] A. H. Raza and K. Nanath, "Predicting a hit song with Machine Learning: Is there an apriori secret formula?," *2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA)*, 2020.

[4] J. Aum, J. Kim, and E. Park, "Can we predict the Billboard Music chart winner? machine learning prediction based on Twitter artist-fan interactions," *Behaviour & Information Technology*, pp. 1–14, 2022.

[5] Rishab Jain Bollywood Songs Dataset. Kaggle. https://www.kaggle.com/datasets/rishabjn/bollywoodsongs

[6] "Spotipy", For the Spotify Web API, there is a little Python module called Spotipy. You have complete access to all of the Spotify platform's music data with Spotipy, https://spotipy.readthedocs.io/en/2.22.1/

[7] "Correlation Heatmap", A correlation heatmap is a visual tool that shows the correlation between many variables as a matrix with different colours. Similar to a colour chart, it depicts how closely related various variables are,.

[8] "Z-Score", A z-score, also known as a standard score, provides information on how much a data point deviates from the mean., https://www.statisticshowto.com/probability-and-statistics/z-score/

[9] "t-SNE plot"A statistical technique called distributed stochastic neighbor embedding (t-SNE) assigns each data point a location on a two- or three-dimensional map in order to visualise high-dimensional data ,https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding#:~:text=t%2Ddistributed%20stochastic%20neighbor%20embedding%20(t%2DSNE)%20is,two%20or%20three%2Ddimensional%20map.

[10] "Precision and recall," Wikipedia, 16-Nov-2022. https://en.wikipedia.org/wiki/Precision_and_recall

[11] "5-fold cross validation", A statistical technique called cross-validation is used to gauge the expertise of machine learning models, https://machinelearningmastery.com/k-fold-cross-validation/

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:15:58 UTC from IEEE Xplore. Restrictions apply.

2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)

Histone Deacetylase 6 as A Putative Target in Alzheimer's Disease Therapeutics

1st Shubham Kumar Shrivastav
Molecular Neuroscience and Functional Genomics Laboratory,
Dept. of Biotechnology
Delhi Technological University
Delhi, India
shri.shubh.7@gmail.com

2nd Pravir Kumar
Molecular Neuroscience and Functional Genomics Laboratory,
Dept. of Biotechnology
Delhi Technological University
Delhi, India
pravirkumar@dtu.ac.in

2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC) | 979-8-3503-2379-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAISC58445.2023.10200676

***Abstract*—Degeneration or loss of brain cells called neurons results in several neurodegenerative disorders (NDDs) and currently, millions of people across the world are suffering from them. The increase in the occurrence of these age-related disorders is witnessed majorly due to lifestyle changes. NDDs harm the overall functioning of the body which can also be fatal to life. Alzheimer's Disease (AD) is the most prevalent in the list of NDDs which affects the memory of patients. Oxidative stress performs a wide role in AD pathology and along with that several proteins and enzymes are responsible for the occurrence of AD. One of them is Histone Deacetylase6 (HDAC6). Normally it is responsible for the release of neurotransmitters but in the diseased condition, it becomes abundant in the nucleus which hinders the function and mechanisms and results in dysfunction of the brain. The deteriorated health of the brain leads to AD. This study, utilizing computational-based molecular docking, presents the binding of multiple types of FDA-approved drugs (Anticancer, Antihypertension, Antidiabetic) with the protein for inhibition. To reduce the activity of HDAC6 as a potential therapy for AD, the AutoDock vina-1.5.7 tool was utilized for virtual screening and prediction of the binding affinity of drugs. Following a simulated screening of several different compounds, it was observed that Bicalutamide had the highest binding affinity, measuring -9 kcal/mol. Bicalutamide is an anticancer drug that may be put through additional pre-clinical and clinical testing to produce medications that are repurposed for the treatment of chronic diseases like AD.**

Keywords*—Neurodegenerative diseases, Alzheimer's Disease, Histone Deacetylase6, drug repurposing approach, molecular docking, Bioinformatics.

I. INTRODUCTION

Neurodegenerative Disorders can be defined as the deposition of specific proteins in the brain and neuronal dysfunction that results in serious medical conditions in a person. NDDs are broadly classified into two types; movement disorders and behavioral disorders. Millions of NDD suffering cases have been recorded; the most prevailing are AD and Parkinson's Disease (PD) [1][2]. Even though all the NDDs present a variety of symptoms but all of them coincide with similar fundamentals like neuronal dysfunction, neuronal death, oxidative stress, ubiquitin abnormalities, neuroinflammation, etc.

Alzheimer's Disease which is the most common and prevailing neurodegenerative disorder is caused by the accumulation of proteins in the brain. Several risk factors are responsible for the occurrence of the disease like age, environment, genetics, lifestyle, and many more. Statistical studies mention that by 2060, approximately 13.9 million Americans will be suffering from AD[3].

According to the reports of the National Institute of Health, approximately 6.2 million people are suffering from Alzheimer's Disease, and over a million are having Parkinson's Disease. numerous registered cases are suffering from one or the other neurodegenerative disorders because of the changing environmental conditions, genetic and epigenetic changes, etc. studies also mention that around 1% population of above 60 years are diagnosed with neurodegenerative disorders [4]. Loads of research are being done but a complete cure for AD is still not discovered. Still, the treatment is based on the symptoms and the condition of a person. In the current era, computational studies are very helpful in the diagnosis, prognosis, and treatment of the disease and several models have been proposed for further clinical studies.

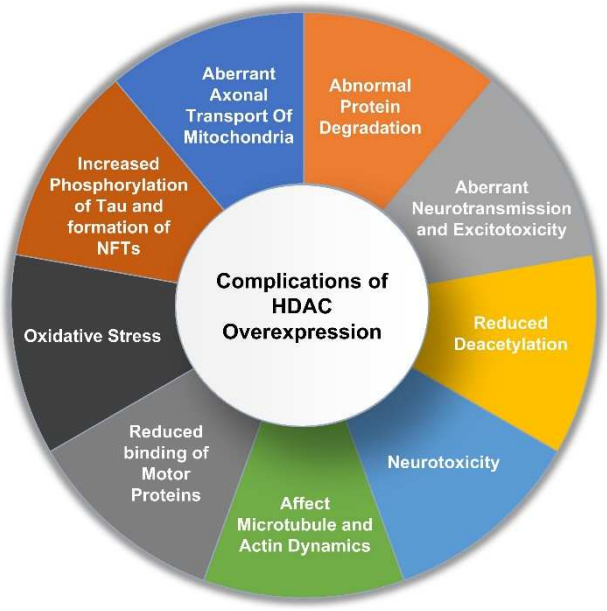


Fig. 1. Cellular complications due to the overexpression of HDAC6

Epigenetic regulation provides a fresh perspective on AD, particularly sporadic AD. Prior research has shown that DNA methylation contributes to AD, and more recently, multiple studies have revealed that histone acetylation contributes to AD pathogenesis [5]. Histone acetyltransferases (HATs) as well as histone deacetylases (HDACs), respectively, catalyze histone acetylation and deacetylation. Chromatin condensation and also gene transcription are significantly influenced by the degree of histone acetylation [6]. According to several studies, HDACs have a crucial involvement in the survival and functioning of neurons. Cellular processes like cell division, development, multiplication, and control of gene expression are dependent on the functioning of HDACs [7]. To carry on such important processes, HDACs are present in 18 different forms in mammals and have been broadly classified into 4 classes based on their structures and

979-8-3503-2379-5/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:49:39 UTC from IEEE Xplore. Restrictions apply.

functions. When HDAC6 forms a compound with polyubiquitinated structures of proteins and shuttles these protein complexes along microtubules to form aggresomes, it accelerates the elimination of misfolded protein clusters [8].

Studies have focused on the genetic factors that may be responsible for the occurrence of the disease and mention that histone deacetylase 6 is worth observing. There is a positive influence of HDAC6 in the synaptic functioning of a healthy brain but in the diseased brain, HDAC6 abundance is highly increased in the nucleus which causes hindrance in the normal functioning of the brain. Over-expression of HDAC6 causes a decrease in α- tubulin which is a microtubule element (Figure 1) [9]. HDAC6 deacetylases proteins like Bruchpilot, tubulin, HSP90, and cortactin. Hyper deacetylation of cortactin and alpha-tubulin can disturb the microtubule and actin dynamics resulting in a compromised neuronal cytoskeleton. Neurotoxic beta-amyloid peptides result in a harmful influx of ions of calcium into brain cells in Alzheimer's disease. It is anticipated that this rise in [Ca2+]int will set off intracellular processes that ultimately lead to cell malfunction and cell death [10]. The increased deacetylation also results in reduced anchoring of motor proteins to alpha-tubulin. Furthermore, deacetylation of peroxiredoxins mediated by HDAC6 will add to oxidative stress contributing to neurodegeneration. The pathogenesis of AD is influenced by oxidative stress, which impairs neuronal mitochondria, oxidizes macromolecules, releases reactive oxygen species (ROS) when metal ions attach to Aβ plaques, and increases p-tau and A production [11]. HDAC6 is present in 4 times the quantity in the AD brain as compared to normal one because of which it can be considered as a source of potent therapy [12]. Finding potent inhibitors for HDAC6 inhibition has been an important step to limit the progression of the disorder (Figure 2). Advancements in the research sector are resulting in several therapies to inhibit the negative consequences that are caused due to the over-expression of HDAC.

In this study, virtual screening of various categories of FDA-approved drugs namely anti-cancer, anti-hypertension, and anti-diabetic passing the BBB was done, and by predicting their binding energy required by the drug to dock at the active site of the HDAC6 and restrict its activity as a therapy for AD. After molecular docking, drugs were sorted based on their binding energy and the root mean square deviation (RMSD) value.

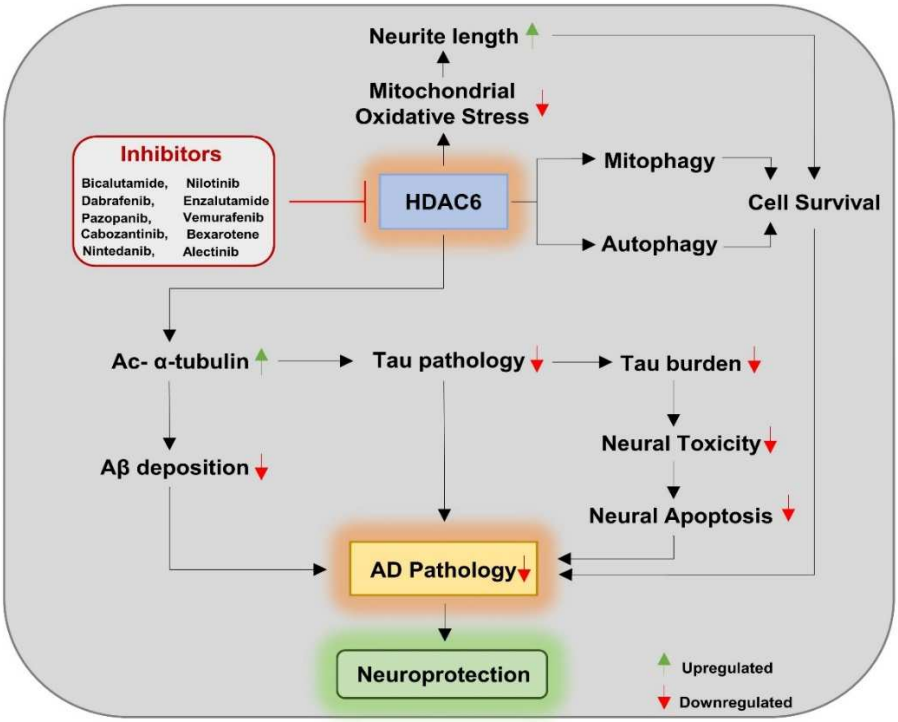


Fig. 2. The reduced activity of the HDAC6 deacetylase, promotes the acetylation of alpha-tubulin, which aids in lowering toxic amyloid segregation, tau pathology, and AB deposition, all of which serve in reduced AD pathology and promote neuroprotection.

A. Structure of HDAC6

HDAC6 belongs to the family of epigenetic modificatory which consists of 18 different kinds of HDACs. All these HDACs are classified into 4 types of HDAC classes. HDAC6 is linked to the class II type. It contains 1215 amino acid residues and is considered the largest member which is present in the zinc finger ubiquitin-binding domain and 2 different catalytic domains (CDs). CD1 is present at the N-terminal and CD2 is present in the central region. The crystal structure for both domains is present which specifies that the CD1 domain shows narrower specificity towards substrate whereas CD2 shows a broader specificity. The zinc finger ubiquitin domain is present at the C-terminal[13].

II. METHODOLOGY

To find the potent inhibitors for the inhibition of HDAC6, computational work is performed which comprised several steps from retrieval of drugs to molecular docking. The potent drugs' major criterion was passing the Blood Brain-barrier Permeability. Here the entire work is based on the drug repurposing approach.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:49:39 UTC from IEEE Xplore. Restrictions apply.

A. Data Collection

Initially, to understand the theoretical concepts of HDAC6 and neurodegeneration, the information was summarized from the literature work presented in literature databases like PubMed https://pubmed.ncbi.nlm.nih.gov/, Google Scholar https://scholar.google.com/. For the study, three classes of FDA-approved drugs were shortlisted based on diseases like anti-cancer, anti-diabetes, and anti-hypertension. As AD is a multifactorial disease, it has been observed that factors like diabetes, hypertension increased oxidative stress and cancer also play a significant role in adding to the progression of AD, on this evidence these classes of drugs were taken in for study as the drugs repurposed in this study will act as multi-target compounds. The database used was PubChem https://pubchem.ncbi.nlm.nih.gov/.

The classification of the drug list was done as 150 anti-cancerous drugs, 50 anti-diabetic drugs, and 50 anti-hypertension drugs. The structure of the receptor (3c5k) was retrieved from the PDB database https://www.rcsb.org/. Finally, the Blood Brain-barrier permeability of the enlisted drugs was predicted using CB Ligand https://www.cbligand.org/CCGS/. The development of drugs for AD is challenged by the blood-brain barrier (BBB), which is also the single most important barrier restricting the development of neurotherapeutics in the future. In light of this, a BBB predictor specifically determines if a substance may pass through the blood-brain barrier (BBB+) or not (BBB-). The support vector machine (SVM) and LiCABEDS algorithms were used to create this predictor by applying them to four different types of fingerprints (MACCS; OpenBabel (F2P); MolPrint 2D; PubChem) from 1593 reported chemical compounds. Among all, SVM algorithm and the PubChem fingerprints were used to predict the compounds for BBB+ or BBB- [14].

B. Receptor and Ligand preparation

For the preparation of the target protein BIOVIA discovery studio is used which is a science-based valid application https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/. Here, water molecules and hetero-atoms were removed, and the energy minimization and polar hydrogen atoms were added using the inbuilt AutoDock vina tool in PyRx. Further, for ligand preparation, all the ligands were converted into a single structure data file (SDF) library. The Open Babel tool integrated with PyRx was used for energy minimization and conversion of ligand files to pdbqt format. The grid mapping was done for HDAC6 and the dimensions were 5.669, -4.987, 12.262 and the size was 34.86 x 39.07 x 40.61.

C. Molecular Docking

The receptor and ligands that were prepared were docked using the AutoDock vina incorporated in the PyRx tool and the top 10 best drugs were selected based on their binding energy and RMSD value i.e., root mean square deviation values. The expected binding energy was to be -9 kcal/mol.

PyRx is a computational tool used to screen the compound libraries against a specific receptor and is an in-silico approach for drug discovery and development. It also comprises of docking wizard and visualization engine used to perform virtual screening and molecular docking. It uses AutoDock Vina as a docking software[15]. AutoDock vina utilizes a scoring function based on empirical force fields to estimate the binding affinity between a ligand and the protein. The scoring function considers various energetic contributions, such as van der Waals interactions, electrostatic interactions, and de-solvation effects, to evaluate the binding affinity. The underlying mathematical model in AutoDock Vina involves algorithms for global optimization, specifically an algorithm called iterated local search. This algorithm performs a systematic exploration of the search space to find the most energetically favorable binding conformations of the ligand within the protein's binding site. Open Babel for the import of files in SDF format. Its programming language is Python and matplotlib is used for 2D plotting.

III. RESULTS

After virtually screening all the retrieved compounds against HDAC6, a list of best-binding drugs was obtained (Table II). Bicalutamide, an anti-cancerous drug, resulted as the novel drug which shows the best binding that can be repurposed to inhibit HDAC6 and limit the progression of Alzheimer's Disease. The ChEMBL ID for the drug is ChEMBL409. The calculated binding affinity of Bicalutamide was -9kcal/mol and the IC50 value was 470 nM. The active site residues of HDAC6 interacting with the bicalutamide are GLU-33, ARG-47, and TYR-81 with a bond length of 3.3 Å, 1.8 Å, and 2.5 Å respectively (Figure 3). This drug passed the test of Blood Brain-Barrier permeability as well and can be considered for further pre-clinical and clinical trials. Promising candidates who can inhibit HDAC6 and slow the course of AD were found using the performance metrics, including binding affinity, docking accuracy, and computational efficiency (Table 1).

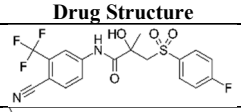
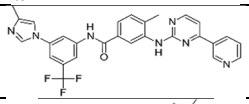
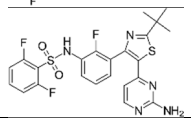
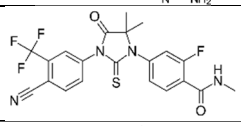
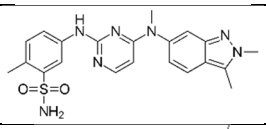
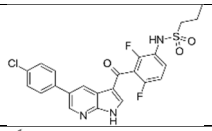
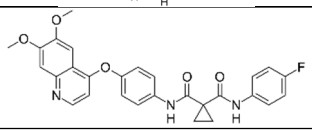
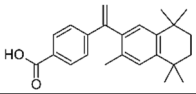
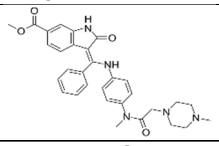
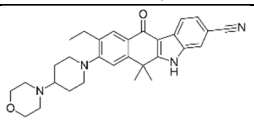
TABLE I. PERFORMANCE METRIC OF AUTODOCK VINA OF THE STUDY CONDUCTED.

Performance Metric	Description	Evaluation Method	Results
Binding Affinity	Measures the strength of ligand-protein interaction	Calculation of binding affinity scores	Average binding affinity (top 10): -8.54 kcal/mol
Docking Accuracy	Evaluates the accuracy of docking predictions	Comparison of predicted binding poses with experimental data	RMSD less than 1.0 Å
Computational Efficiency	Measures the speed and scalability	Calculation of the time required for docking calculations	Average docking time: 2 seconds

Bicalutamide is a drug used for the treatment of prostate cancer. It is an androgen receptor inhibitor. The brand name under which this drug is sold is Casodex. The Drug Bank accession number is DB01128. Bicalutamide is an FDA-approved drug that is always taken in oral form. The trace remains of the drugs are easily excreted through urine and fecal matter[16]. Bicalutamide is a non-steroidal drug but when consumed can cause side effects like weakness, dizziness, swelling in various parts of the body, pain, constipation, nausea, fever, high blood pressure, frequent urination, etc. It should be stored at room temperature and should be consumed only when prescribed by medical experts. Bicalutamide is a multi-purpose drug that can be repurposed to treat neurological disorders like Alzheimer's Disease.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:49:39 UTC from IEEE Xplore. Restrictions apply.

TABLE II. BEST BINDING COMPOUNDS WITH HDAC6 AFTER THE VIRTUAL SCREENING.

Drug Structure	ChEMBL ID	Compound	Binding Affinity (Kcal/Mol)	Blood-Brain Barrier
	CHEMBL409	Bicalutamide	-9	+
	CHEMBL255863	Nilotinib	-8.8	+
	CHEMBL2028663	Dabrafenib	-8.8	+
	CHEMBL1082407	Enzalutamide	-8.6	+
	CHEMBL477772	Pazopanib	-8.6	+
	CHEMBL1229517	Vemurafenib	-8.5	+
	CHEMBL2105717	Cabozantinib	-8.3	+
	CHEMBL1023	Bexarotene	-8.3	+
	CHEMBL502835	Nintedanib	-8.3	+
	CHEMBL1738797	Alectinib	-8.2	+

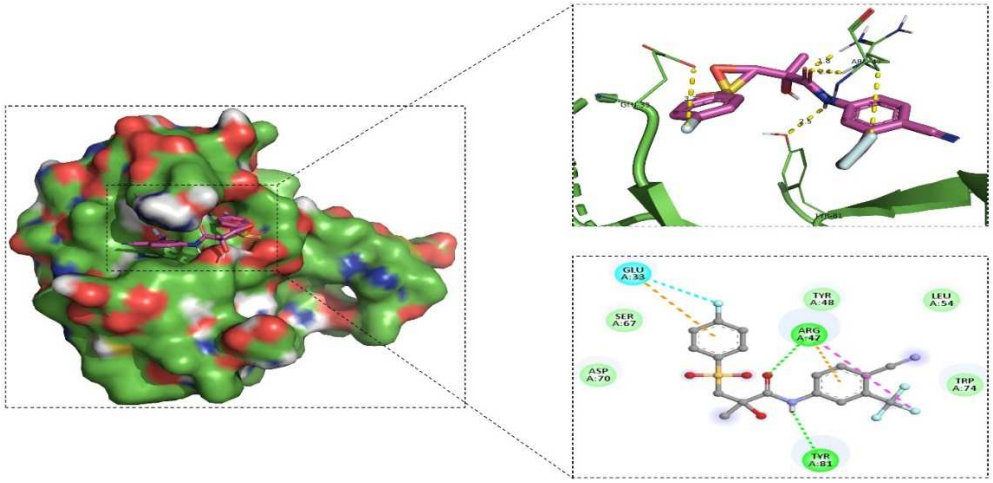


Fig. 3. Binding of Histone Deacetylates 6 (Zinc finger domain) (3c5k) with Bicalutamide, active site interacts with GLU-33, ARG-47, and TYR-81.

IV. DISCUSSION AND CONTRIBUTION

With the induction of dynein and kinesin motor complexes, inhibition of HDAC6 with an inhibitor reverses oxidative stress-induced neuronal cell death and triggers neurite extension, enhances microtubule stability, and mitochondrial transport enhances cognitive and memory conditions associated with diseases like Alzheimer's and neurological impairments both in cell culture and animal

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:49:39 UTC from IEEE Xplore. Restrictions apply.

experimental models. In the R6/2 mouse model, genetically inhibiting HDAC6 has a twofold impact on the development of the illness and motor deficits. This research collectively demonstrated the identification of HDAC6 inhibitors as therapeutics for neurodegenerative disorders [17]. The identified inhibitors were found to be traditionally used for multiple other diseases such as cancer, hypertension, and diabetes.

The points mentioned below highlight the contribution of the Insilco drug repurposing research for the treatment of AD.

- This insilico drug repurposing study finds novel uses for already approved drugs for the inhibition of overexpressed HDAC6 for the treatment of AD by studying the pharmaceutical-target protein molecular interactions.
- Drug repurposing speeds up medication development for neurological disorders like Alzheimer's Disease, Parkinson's Disease, etc. In silico tools can swiftly screen large chemical libraries against many target proteins to uncover potential hits and lead compounds for further investigation. Drug repurposing helps us by saving time and money by using existing data and knowledge.
- Existing drugs that are repurposed for new indications can offer a more comprehensive safety profile than newly developed compounds. Due to the fact that the safety profiles of repurposed pharmaceuticals have already been established through clinical use, the risks associated with adverse effects and toxicity can be better comprehended.
- In silico drug repurposing initiatives provide valuable insights into the action mechanisms and target interactions of existing drugs. This information can aid rational drug design by guiding the modification of existing compounds or the creation of new derivatives with enhanced efficacy or fewer adverse effects. By comprehending the structural and functional properties of repurposed pharmaceuticals, researchers can rationally enhance their therapeutic potential to treat AD.
- Drug repurposing avoids unnecessary animal testing by using approved drugs' safety and knowledge. Repurposed medications have undergone significant preclinical and clinical testing, thus animal testing is unnecessary.

V. Conclusion

FDA-approved drugs which are responsible for the treatment of a variety of diseases were taken for the inhibition of HDAC6 whose over-expression in the brain can cause Alzheimer's disease. Multiple bioinformatics tools and techniques along with online databases were used to conduct the study based on the drug-repurposing approach. Bicalutamide, an anti-cancerous drug resulted to be a potent drug for the inhibition of HDAC6 and curing of AD.

Computational approaches can be one of the best ways to research in the field of life sciences and healthcare which can be cost-effective, and time-saving. At the same time, the only limitation of computational work is its findings need to be approved by clinical trials and further research for human administration. As technology advances, it can be helpful to figure out precise treatments for NDDs for which treatments are yet unknown.

Acknowledgement

We would like to thank the lab members of Molecular Neurosciences and Functional Genomics at the Department of Biotechnology, of Delhi Technological University for their constant support and help in the completion of this project.

References

[1] J.-Y. Liu, H.-Y. Guo, Z.-S. Quan, Q.-K. Shen, H. Cui, and X. Li, "Research progress of natural products and their derivatives against Alzheimer's disease," J Enzyme Inhib Med Chem, vol. 38, no. 1, Dec. 2023, doi: 10.1080/14756366.2023.2171026.

[2] R. Balestrino and A. H. V. Schapira, "Parkinson disease," Eur J Neurol, vol. 27, no. 1, pp. 27–42, Jan. 2020, doi: 10.1111/ene.14108.

[3] K. A. Matthews et al., "Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged ≥65 years," Alzheimer's and Dementia, vol. 15, no. 1, pp. 17–24, Jan. 2019, doi: 10.1016/j.jalz.2018.06.3063.

[4] L. Stickley, R. Koch, and E. Nagoshi, "The utility and caveat of split-GAL4s in the study of neurodegeneration," Fly (Austin), vol. 17, no. 1, Dec. 2023, doi: 10.1080/19336934.2023.2192847.

[5] R. M. Stilling and A. Fischer, "The role of histone acetylation in age-associated memory impairment and Alzheimer's disease," Neurobiol Learn Mem, vol. 96, no. 1, pp. 19–26, Jul. 2011, doi: 10.1016/j.nlm.2011.04.002.

[6] K. Xu, X.-L. Dai, H.-C. Huang, and Z.-F. Jiang, "Targeting HDACs: A Promising Therapy for Alzheimer's Disease," Oxid Med Cell Longev, vol. 2011, pp. 1–5, 2011, doi: 10.1155/2011/143269.

[7] A. De Simone and A. Milelli, "Histone Deacetylase Inhibitors as Multitarget Ligands: New Players in Alzheimer's Disease Drug Discovery?," ChemMedChem, vol. 14, no. 11. John Wiley and Sons Ltd, pp. 1067–1073, Jun. 05, 2019. doi: 10.1002/cmdc.201900174.

[8] C. R. Guthrie and B. C. Kraemer, "Proteasome Inhibition Drives HDAC6-Dependent Recruitment of Tau to Aggresomes," Jwenal of Molecular Neuroscience, vol. 45, no. 1, pp. 32–41, Sep. 2011, doi: 10.1007/s12031-011-9502-x.

[9] Y. Li et al., "Inhibition of Histone Deacetylase 6 (HDAC6) as a therapeutic strategy for Alzheimer's disease: A review (2010–2020)," European Journal of Medicinal Chemistry, vol. 226. Elsevier Masson s.r.l., Dec. 15, 2021. doi: 10.1016/j.ejmech.2021.113874.

[10] B. J. Blanchard, G. Konopka, M. Russell, and V. M. Ingram, "Mechanism and prevention of neurotoxicity caused by β-amyloid peptides: relation to Alzheimer's disease," Brain Res, vol. 776, no. 1–2, pp. 40–50, Nov. 1997, doi: 10.1016/S0006-8993(97)01003-2.

[11] L. Cassidy, F. Fernandez, J. B. Johnson, M. Naiker, A. G. Owoola, and D. A. Broszczak, "Oxidative stress in alzheimer's disease: A review on emergent natural polyphenolic therapeutics," Complement Ther Med, vol. 49, p. 102294, Mar. 2020, doi: 10.1016/j.ctim.2019.102294.

[12] V. Kumar, S. Kundu, A. Singh, and S. Singh, "Understanding the Role of Histone Deacetylase and their Inhibitors in Neurodegenerative Disorders: Current Targets and Future Perspective," Curr Neuropharmacol, vol. 20, no. 1, pp. 158–178, Jun. 2021, doi: 10.2174/1570159x19666210609160017.

[13] Y. Hai and D. W. Christianson, "Histone deacetylase 6 structure and molecular basis of catalysis and inhibition," Nat Chem Biol, vol. 12, no. 9, pp. 741–747, Sep. 2016, doi: 10.1038/nchembio.2134.

[14] H. Liu et al., "AlzPlatform: An Alzheimer's Disease Domain-Specific Chemogenomics Knowledgebase for Polypharmacology and Target Identification Research," J Chem Inf Model, vol. 54, no. 4, pp. 1050–1060, Apr. 2014, doi: 10.1021/ci500004h.

[15] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," J Comput Chem, p. NA-NA, 2009, doi: 10.1002/jcc.21334.

[16] I. D. Cockshott, "Bicalutamide Clinical Pharmacokinetics and Metabolism," 2004.

[17] W. Jian et al., "Inhibition of HDAC6 increases acetylation of peroxiredoxin1/2 and ameliorates 6-OHDA induced dopaminergic injury," Neurosci Lett, vol. 658, pp. 114–120, Sep. 2017, doi: 10.1016/j.neulet.2017.08.029.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:49:39 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the 5th International Conference on Inventive Research in Computing Applications (ICIRCA 2023)
IEEE Xplore Part Number: CFP23N67-ART; ISBN: 979-8-3503-2142-5

Huntington's Disease Biomarker Discovery: A Computational Approach

Jasmeer Singh Kalra
Department of Biotechnology
Delhi Technological University
New Delhi, India – 110042
jasmeersingh_2k20bt509@dtu.ac.in
https://orcid.org/0000-0001-7836-6860

Pravir Kumar
Molecular Neuroscience and Functional Genomics Laboratory
Department of Biotechnology
New Delhi, India – 110042
pravirkumar@dtu.ac.in

2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) | 979-8-3503-2142-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIRCA57980.2023.10220841

Abstract— **Huntington's disease (HD) is a neurodegenerative disorder that can be inherited from a person who has a mutant copy of the HTT gene. This mutated HTT gene encodes the mutated huntingtin (mHTT) protein. This disease targets the basal ganglia and cortex which is responsible for the movement of the body and memory cells, gradually it causes the slowing of brain cells and involuntary movements in the body. Progressively cause deterioration of the vital functions of the body which leads to complications and premature death. To track the disease progression and to alleviate novel therapies identification of the disease is essential. Biomarkers provide reliable effective and accurate prognosis of the disease. Here, the application of computational algorithms and mathematical modeling techniques to identify HD is done. Global Unified Huntington's Disease Rating Scale (UHDRS) traits are used in conjunction with a TRACK-HD dataset of patients with symptomatic and managed Huntington's disease to predict HD. Predictive models for early HD patients, pre-manifest gene carriers, and control patients were created using machine learning and deep learning. These predictors were used as Biomarkers for the prediction of Huntington's disease.**

Keywords—Huntington's disease, Huntingtin (HTT) gene, Basal Ganglia, and Cortex, Biomarkers, Unified Huntington's Disease Rating Scale (UHDRS)

I. INTRODUCTION

Huntington's disease (HD) is a rare genetic neurodegenerative disorder caused by the presence of trinucleotide CAG repeats in the Huntington's (HTT) gene. These repetitive sequences trigger the production of an unstable protein, leading to the formation of inclusion bodies in the brain. These bodies are small toxic fragments that degenerate the functioning of neurons inside the brain because of which brain cells die. Along with these, it causes interference with a number of tasks performed for movement and thought processes. [1] These afterward are responsible for disrupting characteristic motor psychiatric tasks of the brain. As mentioned, it is a rare disease and it affects 1 in every 10,000 people. It usually occurs between the ages of 30-50. Studies have proved that after onset of symptoms average lifespan is about 15 to 25 years of age and over 200,000 persons in the US are thought to be at risk of inheriting HD, and it is believed that about 30,000 of them have symptoms. The World Health Organization (WHO) estimates that HD causes 0.4% of global fatalities. [2] Unfortunately, medical research can neither halt it from becoming worse nor find a solution. Hence, the identification and prediction of this disease can have a significant impact as it can further help us develop personalized therapies. Biomarker discovery can be used for identifying novel drug targets used for drug discoveries and for the prognosis of the disease. Computational Biologists can manage to help patients by predicting disease and planning treatment accordingly as early treatment can reduce side effects and progressive disease will prevent fortunate treatment. [1]

TRACK-HD is a prospective Huntington's disease biomarker research. It employs many investigations to diagnose HD, including multiple MRI and image-based scans, as well as blood-based biomarkers, and it assesses the development of HD. Many of these were changes in pre-symptomatic behaviour, while others were changes that occur over a period of time. [4] The Unified Huntington's Disease Rating Scale, also known as UHDRS, is a tool that allows us to measure how severe HD has been over time. UHDRS uses a number of features like cognitive function, motor function, and diagnosis confidence to evaluate the efficacy of treatment for HD. In the dataset available in the paper, usage of oculomotor abnormalities of early and pre-manifest HD is done. [3] The study used both TRACK-HD and UHDRS data to fit mathematical and computational models to the publically available to differentiate HD before symptom onset and controlled HD patients.

II. SIGNIFICANCE

Huntington's disease can be characterized by symptoms related to psychiatric and cognitive health. Biomarkers derived from antisaccade testing data can help in the early diagnosis and detection of the disease enabling treatments in earlier stages. The features used to provide measurements of eye movements such as duration, amplitude, latency, and peak velocity can be a quantifiable biomarker for oculomotor function associated with HD. The accuracy and reduced reliance provide subjective assessments. Changes in eye movements provide insights into the development and progression of cognitive impairments which can further help in personalized treatment plans for therapeutic interventions. [3]

By analyzing the saccadometer data alongside the clinical and demographic information, researchers can explore associations between eye movement abnormalities that

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:13 UTC from IEEE Xplore. Restrictions apply.

contribute to HD and its therapeutic targets. The identification of distinct classified clusters within HD using saccadometer data can aid in disease subtyping and stratification. It also may help in predicting progression, tailoring interventions, and assessing treatment in different subgroups. Hence, saccadometer-derived biomarkers can be employed for the evaluation of HD. They provide objective and quantifiable metrics to assess the effectiveness of therapeutic interventions, potentially facilitating the development of new treatments and interventions. [3]

The data generated from different clinical assessments, imaging data, cognitive tests, and biological samples contributed various insights into the natural history of HD and its biomarker discovery through the TRACK-HD study. Whereas, UHDRS is widely used in clinical trials used to assess the severity of disease and track it over a while. This data can also evaluate the effectiveness of biomarker and therapeutic interventions. [8] [9]

III. SACCADOMETER TESTING

Oculomotor testing of the subjects was done using a saccadometer advanced machine that tracks eye movements. In this method, after the saccadometer is put on their head, on a computer screen two lights appear in the center (green and red). The subjects are instructed to follow or take away their eyes from the target as informed (green for prosaccade and red for antisaccade). This data provides accuracy and velocity of saccades to differentiate between healthy and HD patients. It also provides information relevant data about the cognitive functions in HD i.e. cognitive flexibility and inhibition. While checking the error rate in the patients differentiating from that of HD it was negligibly less (less than 1%). An illustration of the same has been shown below. [10]

Oculomotor abnormalities are commonly observed in HD patients, even in the early stages of the disease. By assessing eye movement parameters, such as saccades, smooth pursuit, and fixation stability, oculomotor testing can provide objective measurements of oculomotor function. These measurements can be compared across different populations, disease stages, or treatment responses to identify patterns or deviations specific to HD. Quantitative analysis of oculomotor data may reveal distinct oculomotor biomarkers associated with the disease. These biomarkers can aid in the early detection and diagnosis of HD, differentiate HD from other conditions with similar symptoms, monitor disease progression, and potentially predict disease outcomes. Furthermore, understanding the mechanistic relationship between oculomotor abnormalities and underlying neural dysfunction in HD can provide insights into the disease's pathophysiology and guide the development of targeted therapeutic interventions. Integrating oculomotor assessments with other clinical and imaging biomarkers may enhance the accuracy and efficacy of HD diagnosis, management, and treatment.

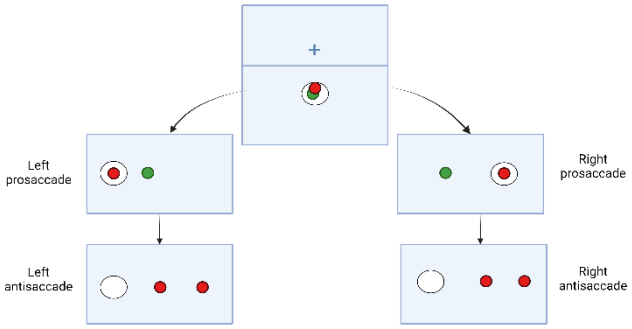


Figure 1: Antisaccade and Prosaccade task performed by saccadometer
Created with BioRender.com.

IV. METHODOLOGY

A dataset containing tests of healthy patients, symptomatic patients, and early-HD patients was taken into account and an Antisaccade testing using a acidometer was done. [5] Antisaccade are determined eye movements in which a person must consciously avoid producing a reflexive eye movement in the direction of an object to produce a rapid eye movement in the opposite direction. This test may be utilized as a diagnostic tool since it can identify irregularities in antisaccade performance in HD related to cognitive and motor function impairments. According to research, people with HD may fail to regulate reflexive saccades, which may result in more mistakes and shorter reaction times during antisaccade activities than in healthy people.

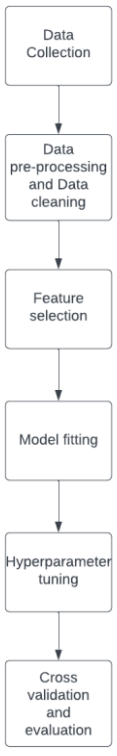


Figure 2: Methodology Block Diagram
Created in Lucidchart, www.lucidchart.com.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:13 UTC from IEEE Xplore. Restrictions apply.

An In-silico analysis of data was collected and examined over multiple countries. A total number of 371 subjects were part of this examination, with a total number of 90476 entries (approximately 250 trials per subject). These individuals underwent an antisaccade test in which they had to move away from an incoming stimulus. These subjects were divided into 1. Control 2. Pre-HD Gene Carrier 3. Early HD with 123 healthy subjects, 122 pre-HD, and 125 HD patients. The groups were further divided into five sub-groups for classifying the progression of HD including control, pre-HDA, pre-HD-B, HD-1, and HD-2. [5]

UHDRS features were used for the evaluation of HD symptom severity, specifically TFC or Total functional capacity and TMS or Total motor scores. Some important features taken into account with the model fitting were –

A. CAG repeats

The HTT or Huntingtin gene contains CAG sequences (cytosine, adenine, and guanine) with several repeated sequences. In an HD diagnosed patient repeats are recorded to no less than 36. This abnormal expansion of the CAG sequence is responsible for an unstable huntingtin protein called mHTT. Which further leads to neuron malfunctioning and progressively death. CAG determination is an important factor for considering whether a person is symptomatic of HD or not. As per studies, individuals with a CAG sequence count of more than 40 experience progressive symptoms of HD. The mean CAG recorded in pre-manifest gene carrier (mean 43.16) and early HD (mean 43.57) patients are approximately the same with a difference of less than 1%. The main reason behind using the CAG repeats data is, it comes under UHDRS features for detection of HD. [11]

B. Age

Age plays an essential role in the progression of Huntington's disease (HD), as mentioned previously CAG repeats are responsible for the mutation of Huntingtin protein which happens at an earlier stage. Individuals at younger age experience faster disease progression and more severe symptoms and rapid increase. Whereas, on the other hand, at a later age onset it is believed that the symptoms will tend to be milder as compared to that of younger persons. Therefore, age plays an important role in the prognostic research of a patient.

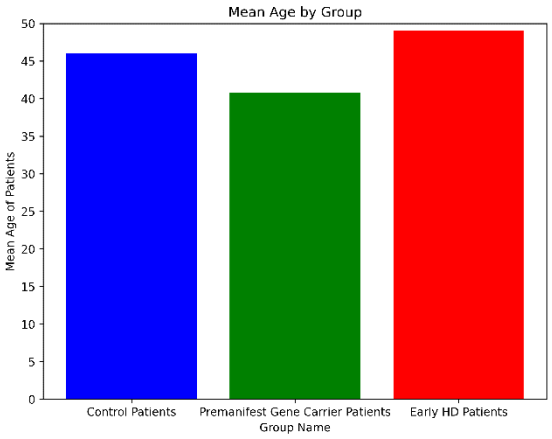


Figure 3. The average age of patients from all groups.

C. Duration, Amplitude, and Foreperiod

Duration is the amount of time it takes a person to finish a saccade, which is a quick eye movement required to focus on a specific object. The length of saccades has been shown to lengthen in HD patients, which may reflect a loss in the brain's capacity to interpret information pertaining to eye movement. The amplitude of a saccade is its angular separation. Saccadic amplitudes in HD patients can vary, and some evidence suggests that amplitudes decrease as the illness progresses. The foreperiod is the amount of time that passes between the appearance of a target and the first warning signal. People with HD had longer foreperiods than healthy controls, according to studies, which may indicate a weakness in their ability to plan an upcoming action. According to our observation, the mean duration time for control patients was 51.86ms, the mean duration time for pre-manifest gene carriers was 53.32ms and the mean duration time for early-HD patients was 58.46ms. There was a slight difference between all three groups by which it can be assumed that duration, amplitude, and foreperiod act as a biomarker for HD.

D. Reaction time, Latency, and Correctness

Reaction time, Latency, and Correctness account for the prognosis of HD, studies have proved that patients suffering from HD tend to have longer reaction time and latency and poor accuracy as compared to healthy individuals. These defects originate from dysfunction in the basal ganglia, a brain region that is affected by the disease. These HD patients perform poorly in these cognitive tasks because of an uncertain involuntary movement in the body. The mean reaction time for control patients was 324.37ms, the mean reaction time for pre-manifest gene carrier was 330.10ms and the mean reaction time for early-HD patients was 366.14ms. It can be witnessed that the reaction time for early HD is slightly higher than that of control patients and pre-manifest gene carrier patients.

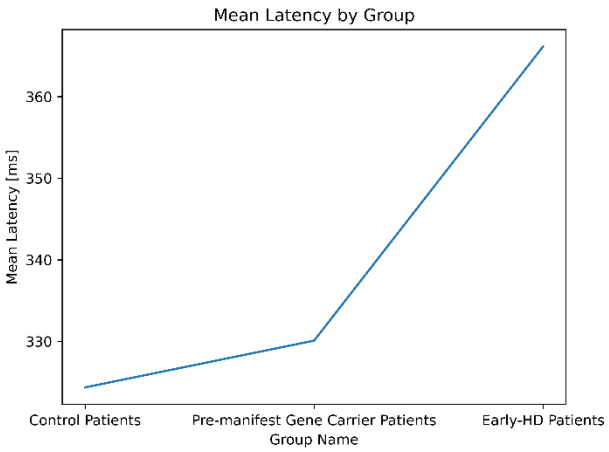


Figure 4. Average Latency [in ms] of patients from a different group.

Whereas for reaction time and correctness, no patterns or correlations were found. It came out to be similar for all the groups. These are essential factors for UHDRS and are hence used in the paper.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:13 UTC from IEEE Xplore. Restrictions apply.

E. TFC, TMS, and DCS

Clinical measures which account for prognosis and intensity measurement of HD are TFC or Total Functionality Capacity, TMS or Total Motor Score, and DCS or Diagnostic Confidence Level Score. These measures are used in the computational modelling technique to differentiate between different subgroups of HD. [12]

A clinical grading system called Total Functional Capacity assesses the functional capacities of people suffering from HD. There are 5 stages in total, where Stage 1 shows full functional ability and Stage 5 denotes total loss of independence. A clinical rating measure called Total Motor Score is used to evaluate motor symptoms in HD patients. A clinical grading system called the Diagnostic Confidence Level Score assesses a clinician's level of assurance in making the HD diagnosis. Higher scores on the 5-point scale indicate greater confidence in the diagnosis. The TFC, TMS, and DCS scores are used in the UHDRS (Unified Huntington's Disease Rating Scale) and TRACK-HD studies to measure many aspects of Huntington's disease. After applying statistical methods to these values it was observed that the TMS score in each group varied from 1.46 in Control Patients, to 2.51 in Pre-manifest Gene Carrier, to 22.92 in Early-HD patients. [13]

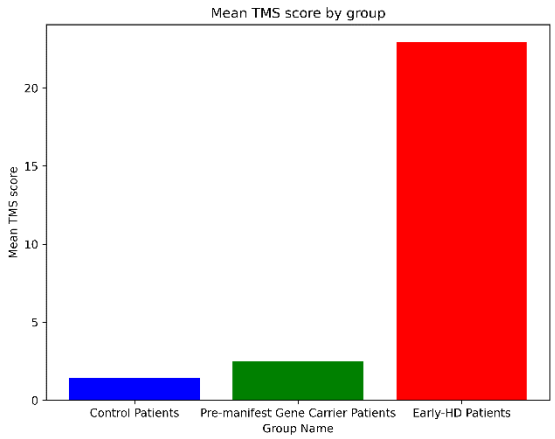


Figure 5: Average Total Motor Score (TMS) for different groups of patients

V. Machine Learning and Deep Learning

The main aim of this research is to assess the accuracy of tests used to diagnose HD in real-world circumstances utilizing computational models like Machine Learning and Deep Learning. The goal was to create a model that could categorize individuals into different phases of HD or forecast the rate of disease development. Before analyzing the TRACK-HD dataset, data cleaning and pre-processing were carried out. Rows with missing data were removed, providing a clean CSV file.

The study employed a dataset of patients with HD and collected pertinent clinical and demographic data to fulfil the research aim. After then, the dataset was divided into training and testing sets. The study determined the relevant features and target variable, then trained a CatBoost classifier on the training set with the specified features and target variable. [7] Accuracy and AUC-ROC scores were collected as performance metrics, and the trained model was then used to generate predictions on the test data. The main purpose was to develop classifiers based on behavioral and model features in a sample of persons with HD, and then investigate if these classifiers could distinguish between new subject groups. Investigation of alternative deep learning and machine learning approaches was done before employing CatBoost, including Long Short-Term Memory (LSTM), Linear Discriminant Analysis (LDA), Random Forest, Gradient Boosting Machine (GBM), Support Vector Machines (SVM), Logistic Regression, and K-Nearest Neighbours (KNN). To examine how effectively a classifier might predict the phases of HD or the rate at which the sickness will advance, Area Under the Receiver-Operator-Characteristic Curve (AUC), a measure robust to unequal class sizes was evaluated. As per our examination, the results determined that CatBoost was the most appropriate algorithm for our dataset. [14] [15]

TABLE I. MODEL AND ITS ACCURACY COMPARISON.

Assessment of Machine Learning Models	
Model Name	***Accuracy***
Support Vector Machine	0.72977
Gradient Boosting Machine	0.94530
Random Forest	0.99665
CatBoost	0.99888
Long Short-Term Memory	0.45358
Linear Discriminant Analysis	0.66853
Logistic Regression	0.75728
K-Nearest Neighbors	0.66095
Gaussian Naive Bayes	0.69609

VI. Validation

A prominent statistic for examining binary classification challenges is the Receiver Operating Characteristic (ROC) curve. In this study, ROC curve analysis was done to determine the accuracy of our Catboost model in predicting the incl02 score for HD. The model demonstrated a high degree of accuracy, with an AUC-ROC of 0.9999, indicating that it can accurately distinguish between the control, premanifest gene carrier, and early HD groups. Furthermore, the precision of the model was determined to be 0.9988, indicating that the model's predictions are highly accurate. These findings indicate that the features extracted from the UHDRS and TRACK-HD datasets have the potential to be biomarkers for the early detection of HD, and they suggest that the Catboost algorithm is a promising tool for biomarker discovery in this field. Overall, our findings show that machine learning technologies have the potential to develop effective biomarkers for early identification and treatment of HD. [6] [15]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:13 UTC from IEEE Xplore. Restrictions apply.

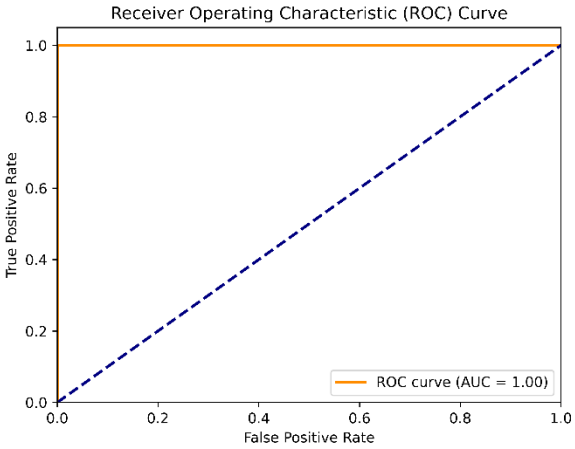


Figure 6: ROC curve of CatBoost Method.

VII. Conclusion

A Machine Learning Model was fitted into a TRACK-HD dataset with UHDRS features in an antisaccade testing that will help clinicians to accurately classify individuals as either control, premanifest gene carrier, or early HD based on certain features or biomarkers present in their data. The dataset used to train the model created in this dataset contains characteristics linked to oculomotor impairments in premanifest and early HD patients. These characteristics are used by the computational model to forecast a person's chance of developing HD. The model is particularly effective at accurately identifying people with HD since it has a high accuracy rate of more than 99%.

This model is thus suitable as a biomarker for HD, as clinicians and researchers can identify people who are at risk for HD before any symptoms appear by using the characteristics that have been found to predict HD. Early detection can enable earlier therapy and management, which can assist decrease the disease's course and enhance patient outcomes [6]. The high performance of more than 99% accuracy is achieved while using CatBoost and Random Forest methods. To further improve the performance, methods such as feature engineering, improvements in data size, data quality, preprocessing, hyperparameter tuning, model ensemble, and the utilization of different cross-validation and evaluation metrics are essential.

In the literature, several similar papers have been published on this topic. However, it should be noted that these papers do not provide any validation scores or accuracy measurements, making it difficult to assess the reliability and performance of their proposed methods. Furthermore, the correlation between different features and patients has not been extensively studied or reported in these papers. Therefore, there is a need for further research to explore the relationship between the different features and their impact on patient outcomes, as well as to validate the proposed methods with appropriate evaluation metrics and validation scores. In reviewing existing literature, it becomes apparent that previous studies have explored machine learning algorithms for HD detection, but few have specifically investigated the correlation between different features and HD diagnosis. In our research, which incorporates CatBoost, addresses the gap by demonstrating promising results in accurately identifying HD based on the correlation of various features within the TRACK-HD and UHDRS datasets. This emphasizes the significance of my model in advancing HD diagnosis and potentially enhancing the understanding and management of the disease.

Acknowledgment

We would like to thank the senior management of Delhi Technological University (DTU) for their consistent support and advice with constant supervision.

Data Availability

Both data and jupyter notebook are freely available for access to all under fair use on GitHub.

https://github.com/Jasmeer57/TRACKHD_DATA_Computational_Analysis.git.

References

[1] Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell, 72(6), 971-983.

[2] WHO. (2016). Neurological disorders: Public health challenges. World Health Organization.

[3] Tabrizi, S. J., Reilmann, R., Roos, R. A., Durr, A., Leavitt, B., Owen, G., ... & Johnson, H. (2012). Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. The Lancet Neurology, 11(1), 42-53.

[4] Tabrizi SJ, Langbehn DR, Leavitt BR, Roos RA, Durr A, Craufurd D, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. Lancet neurology. 2009. September;8(9):791–801.

[5] Wiecki TV, Antoniades CA, Stevenson A, Kennard C, Borowsky B, Owen G, Leavitt B, Roos R, Durr A, Tabrizi SJ, Frank MJ. A Computational Cognitive Biomarker for Early-Stage Huntington's Disease. PLoS One. 2016 Feb 12;11(2):e0148409. doi: 10.1371/journal.pone.0148409. PMID: 26872129; PMCID: PMC4752511.

[6] Paulsen, J. S., Langbehn, D. R., Stout, J. C., Aylward, E., Ross, C. A., Nance, M., ... & Hayden, M. (2008). Detection of Huntington's disease decades before diagnosis: the Predict-HD study. Journal of neurology, neurosurgery & psychiatry, 79(8), 874-880.

[7] Fabbri, D., Fava, M., & Manzoni, S. (2020). Long short-term memory neural networks for traffic speed prediction.

[8] Scahill, R. I., et al. (2013). Clinical impairment in premanifest and early Huntington's disease is associated with regionally specific atrophy. Human Brain Mapping, 34(3), 519-529.

[9] Rosenblatt, A., Kumar, B. V., Mo, A., Welsh, C. S., Margolis, R. L., Ross, C. A., & Nance, M. A. (2012). Age, CAG repeat length, and clinical progression in Huntington's disease. Movement Disorders, 27(3), 272-276.

[10] Sharpe, J. A., & Leigh, R. J. (2002). Saccadic eye movements in the diagnosis of neurologic disorders. Current Opinion in Ophthalmology, 13(6), 346-351.

[11] Aziz, N. A., et al. (2009). Weight loss in Huntington disease increases with higher CAG repeat number. Neurology, 73(1), 96-101.

[12] Ho, A. K., et al. (2009). A 4-year study of the natural history of Huntington's disease: The CARE-HD study. Neurology, 73(18), 14691475.

[13] Huntington Study Group. (1996). Unified Huntington's Disease Rating Scale: Reliability and consistency. Movement Disorders, 11(2), 136142.

[14] Prokhorenkova, L., et al. (2018). CatBoost: unbiased boosting with categorical features. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS), 6638-6648.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7), 1145-1159.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:45:13 UTC from IEEE Xplore. Restrictions apply.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-15508-5

Hybrid reversible watermarking algorithm using histogram shifting and pairwise prediction error expansion

Lavi Tanwar[1] · **Jeebananda Panda**[1]

Received: 27 September 2021 / Revised: 26 November 2022 / Accepted: 18 April 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
The exact retrieval of both i.e., watermark and host data using reversible watermarking makes it more suitable for applications like medical, military etc. The techniques which are considered most efficient and are being most widely used to implement it (reversible watermarking) are Prediction Error Expansion and Histogram Shifting as they offer higher embedding capacity with lesser distortion. The methodology proposed in this work is a hybrid one that utilizes the best of both. The algorithm starts with splitting the image into two regions and further processing is done on them one by one thereby covering each pixel for embedding thus amounting to enhanced capacity. The proposed work exploits the histogram of prediction errors instead of difference of adjacent pixels as it is more sharply distributed. This leads to lesser distortion in watermarked image. In this paper, the sorting of prediction errors is done in accordance with the calculated variance values of their prediction context followed by pairwise payload embedding. The results have substantiated the fact that for the smaller variance, prediction errors are also small. The proposed methodology is tested against the standard test images and real life applications such as medical, biomedical, aerial, military and colour images of Kodak image dataset. The experimental results demonstrate the superiority of the proposed work over the existing ones in terms of embedding capacity and distortion.

Keywords Prediction error expansion · Histogram shifting · Steganography · Reversible watermarking · Hybrid predictor

✉ Lavi Tanwar
lavi.tanwar@dtu.ac.in

Jeebananda Panda
jpanda@dce.ac.in

[1] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, India

1 Introduction

The advancement of Internet technology has made the communication so facile that even a person miles away can connect with others without any hindrance. One can effortlessly access any sort of digital data, document, image or video, for any purpose, either for gathering knowledge or for entertainment, and can share it with others. This easy access to the data, at the same time, poses a threat to its security and the owner's copyright claim. One can easily infringe the data without giving the due credit to the owner and can use it unethically. Such fraudulent acts necessitate the requirement for data hiding techniques to avert the misuse of data. The research work steering in this direction suggests Steganography and Digital Watermarking as the optimal solutions for this.

Steganography and Digital Watermarking both insert the covert information within the original data. Steganography focuses only on the secret communication between the sender and the receiver. The unauthorised user is unaware of the presence of secret data [1]. This way the secured communication takes place. On the contrary, the secret information, known as watermark, is used to prove the authenticity of the data. A randomly generated sequence of bits or any image or any information revealing the owner's identity, like owner's name or a company logo, can be used as a watermark to hide within the host data [2, 5, 7]. The researchers have been striving for ameliorating the performance of digital watermarking algorithms in an attempt to preserve the digital content and the rights of its owner, even after its dissemination [3, 4]. Apart from copyright protection, Digital Watermarking also suits to other applications such as fingerprinting, tamper detection, protection of data from being copied and so on [29]. The algorithms so designed may utilize key while embedding the secret information. The same key must be provided during the extraction phase for authentic recovery of hidden data. But sharing of key with the receiver is difficult and moreover, if unauthorized user gets the access of key then the data can be tampered or misused [22]. Thus, the proposed work is blind in nature which embeds the information required for extraction within the cover data itself. This enhances the security of the data while transmission.

In conventional watermarking, original data might get destroyed. However, some applications do exist where even the small changes in the original data cannot be afforded. Healthcare services is one of the major applications. Suppose a patient's MRI is watermarked with his details then on the receiver side the patient's details are required for authentication purpose and his MRI is required for further treatment. This advocates for the need of exact retrieval of both the watermark and the cover data. Military services is another critical application. Both the information content as well as watermark of these applications are very sensitive. Even the small changes made by illegitimate user should be recognised by the receiver. This demands for the fragile technique so that integrity can be assured. To accomplish such objectives, Reversible Watermarking or Lossless Data Hiding algorithms can be utilized which are fragile in nature.

The concept of reversible watermarking, pioneered by Hosinger et al. [12], aims to execute distortionless embedding and recovery of cover image in addition to the watermark. The realization of the same demands for the insertion of supplementary data, required later for extraction, besides the watermark [16, 26, 28]. This however increases the size of the total payload to be embedded as against the conventional one. This stipulates for the need of an algorithm with higher embedding capacity which may, in turn, lead to higher distortion. At the same time the watermark must be imperceptible. Thus, an efficient reversible watermarking algorithm attempts to sustain the balance between all these parameters. The researchers have designed various state-of-the-art techniques to attain these objectives.

The techniques can be broadly classified into three major categories [16, 24]. These are:

1. Histogram modification based techniques
2. Compression based techniques
3. Expansion based techniques

In Histogram modification based techniques, the image histogram bins assist in locating the appropriate locations for hiding the watermark. The algorithms based on this technique are generally semi-fragile in nature. Ni et al. [25] proposed an algorithm utilizing the pixel values within the range of peak and zero points of the cover image histogram for watermark embedding. This showed that the number of pixels in the selected bins constrained the embedding capacity. An attempt has been made by Lin et al. [21] to improve the embedding capacity by utilizing the difference of adjacent pixels of host image to create a difference image. They divided the image into different blocks and applied the Histogram modification on the corresponding blocks of difference image to insert the payload. Although this technique enhanced the embedding capacity but it suffered from increased overhead data as peak value information needed to be stored for each block.

The compression based techniques, before embedding the watermark, compress some part(s) of the host data resulting in the creation of space for data hiding. The total payload, hence, is the combination of both the original watermark and the information related to the compressed part. These algorithms are mostly fragile in nature. In [6], Celik et al. attempted to utilize both the quantization and compression to generate the watermarked image. Memon et al. [23] presented a compression based approach applying threshold optimization that ultimately increased the embedding capacity.

The techniques which are being used extensively today by the researchers come under the category of Expansion based techniques. This is due to the fact that these techniques exhibit better embedding capacity and are computationally efficient. These techniques can further be classified into three categories:

1. Difference Expansion based techniques
2. Prediction Error Expansion based techniques
3. Interpolation based techniques

Difference Expansion (DE) based techniques exploit the redundancy available in the cover image to achieve the higher embedding capacity [35]. These techniques are known for their high embedding capacity and low computational complexity.

Prediction Error Expansion (PEE) based techniques are the extension of DE based techniques which further improves the embedding capacity and reduces the distortion. These predict the values of the candidate pixel using some prediction algorithm by exploiting the correlation between the neighbourhood pixels. To keep the distortion minimum, the prediction algorithm must be capable of predicting accurate value such that negligible prediction error occurs. The commonly used predictors are Mean predictor, Rhombus predictor, Median Edge Detector (MED) predictor. The predicted values are expanded to hide the payload and the expandable locations are determined by the Histogram Shifting (HS) method. Here the histogram of prediction errors is used to determine the locations because it is more sharply distributed leading to better performance. But, the expansion amounts to the "data overflow" and "data underflow" problem. Preferably, the pixels with smaller prediction errors are modified to achieve low distortion [10, 28, 41].

Interpolation based techniques intent to estimate the missing pixels to generate interpolated image. The interpolation errors, thus obtained, are considered for the data hiding. Unlike other DE techniques, for watermark embedding it considers interpolation error rather

than difference between adjacent pixels or prediction errors [15, 32]. These techniques are generally fragile in nature.

Table 1 summarizes the properties of different reversible watermarking techniques. It can be concluded that different techniques serve different purposes, so today researchers prefer to use combination of more than one technique, known as Hybrid technique, to incorporate the merits of various techniques. As it can be seen that PEE techniques provide better embedding capacity and lower computational complexity and HS offers for the expandable locations, so this paper has combined the HS and PEE techniques to design a reversible watermarking algorithm. Moreover, to reduce the distortion, the pairwise expansion of the prediction errors has been done in this paper. The paper also targets to make the proposed algorithm blind.

The structure of the paper is organised as: Section 2 summarizes some of the existing methodologies developed for the reversible watermarking. The Section 3 explains the proposed methodology in detail. In Section 4, the results obtained by implementing the proposed algorithm have been presented and compared with the existing state-of-the-art algorithms. The Section 5 concludes the work presented in this paper.

2 Related works

This section gives insights into the work that has been done in the recent times using HS and PEE techniques. The development of such reversible algorithms start with the determination of the candidate pixels for prediction. These pixels are then predicted using the neighbouring pixels by executing a certain predictor. The predicted values are then used to calculate the prediction errors by taking the difference between them and the original pixels.

$$e = f - f' \tag{1}$$

where 'e' is the prediction error, 'f' is the original image pixel and 'f′' is the predicted value of the pixel f.

Next the Prediction Error Histogram (PEH) of all the prediction errors is derived and modified to determine the expandable locations. This was done in 2007 for the first time by Thodi and Rodriguez in [34]. The researchers applied different mechanisms to modify the PEH that ultimately decided the performance of the designed algorithm. Some algorithms

Table 1 Comparison of different reversible watermarking techniques [16]

Technique	Properties	Remarks
Histogram modification based techniques	Bins of Image histogram are used to identify the locations for hiding the watermark	Embedding capacity is constrained by the number of pixels available in the selected bins.
Compression based techniques	Compresses some part of the host data to create space for data hiding	Mostly fragile in nature and it increases the size of payload to be embedded
Expansion based techniques	Exploits the correlation among the neighbourhood pixels	Higher embedding capacity and it can suffer from data overflow and data underflow problem

select the zero bins to expand the errors while some select the peak bins. On the other hand, some researchers have attempted to select a range for expanding the errors and the bins outside the range are shifted to prevent the overlapping. Figure 1 shows a typical example of a PEH. The errors within the bin range [−1,1] shown by grey colour are expanded while others are shifted. The distortion can further be reduced if the prediction errors are first sorted before embedding the watermark bits. This has been demonstrated by Sachnev et al. [31] where the author tends to first sort the prediction errors on the basis of the local variance.

Later the researchers have devised the pairwise PEE technique so that the redundancy available in the image can be exploited in a better way to further enhance the embedding capacity and reduce the distortion. Ou et al. [27] have exploited the correlation between the adjacent prediction errors and forms the prediction error pairs while Dragoi et al. [8] opted for adaptive strategy for pairing. The work proposed in [40] and [17] has also shown that the algorithms based on pairwise PEE techniques reduce the distortion and provide watermarked images of higher visual quality. The predictors, that are used to predict the values of the candidate pixels, also play a crucial role in reducing the distortion. The accuracy of the predicted values depends on the type of predictor used. The commonly used predictors are Mean predictor, Rhombus predictor, Median Edge Detector (MED) predictor. M. Ishtiaq et al. [14], devised a hybrid predictor to obtain small prediction errors. These advancements have resulted in the improvement in terms of distortion and embedding capacity as compared to the other reversible watermarking techniques. But the results have also shown that with the increase in number of watermark bits, the visual quality of the watermarked images gets deteriorated. This shows that the embedding capacity of these techniques is limited. This calls for the designing of such algorithms that can achieve higher embedding capacity without the significant increase in distortion.

The proposed work is also inspired from the existing works and aims to ameliorate the performance of the existing algorithms and overcome their shortcomings. The contribution of the proposed work can be outlined as:

- The paper targets to accomplish higher embedding capacity with lesser distortion using HS and pairwise PEE techniques. It splits the host image into two regions which subsequently entitle every pixel to accommodate watermark bit leading to higher capacity. The histogram of prediction errors has been devised as opposed to histogram of difference between the adjacent pixels as the former is more sharply distributed leading to lesser distortion.

Fig. 1 Example of Prediction Error Histogram that attempts to select a range [-1,1] shown by grey colour for expanding the prediction errors and the bins outside the range are shifted to prevent the overlapping

- This paper effectuates sorting of prediction errors with the utilization of variance of prediction context and while embedding it forms the pair between the pixels corresponding to the adjacent sorted prediction errors unlike the existing techniques which considers the adjacent pixels for pairing. This realizes the goal of high quality watermarked image. Moreover, it spreads the watermark all over the cover image which makes it difficult to be identified by the unauthorised person.
- The proposed approach has been made a blind one so that no information needs to be sent along the watermarked image. All the prerequisites for the detection algorithm have been inserted in the watermarked image itself.
- The proposed algorithm has been executed on the medical and military images also along with the standard test images to examine its performance on real life applications. The algorithm has been subjected to colour images also.

3 Proposed methodology

3.1 Watermark embedding

The methodology deploys HS and pairwise PEE techniques to realize a blind reversible watermarking algorithm. The procedure for embedding the watermark in the cover image to obtain the watermarked image is shown in Fig. 2. The figure is discussed in detail through the following sub-sections..

3.1.1 Image decomposition

Image disintegration is the prime requirement towards the accomplishment of the desired objective. The image is decomposed into two regions as represented in Fig. 3. This has been done to circumvent the superimposition of prediction context required for predicting the value of the entitled pixel. The superimposition makes it difficult to recover the original image while decoding. As depicted in Fig. 3, the first region is comprised by the pixels represented by '+' sign while the second region is comprised by the

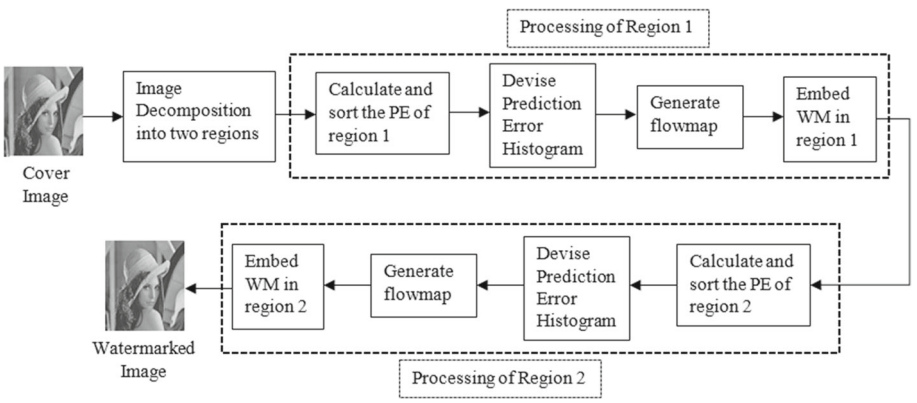


Fig. 2 Procedure to construct watermarked image

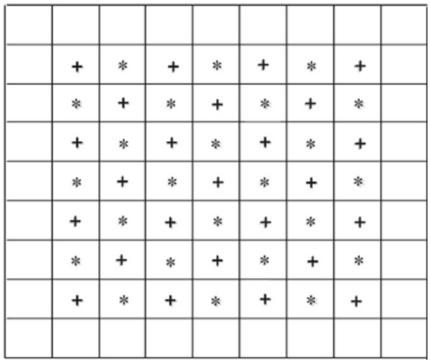


Fig. 3 Decomposition of the cover image into two non-overlapping regions. Region one is marked with '+' sign and Region 2 is marked with '*' sign

ones represented by '*' sign. The peripheral pixels of the cover image are left unaffected due to the requirement of first and last row for the insertion of payload header at the later stage. This makes all the pixels accessible for the embedding purpose except those at the boundaries and thus helps in achieving the objective of high embedding capacity.

3.1.2 Prediction methodology

The next step calls for the calculation of the prediction values of the candidate pixels. The pixels marked with '+' sign are designated as the candidate pixels for the first region and similarly '*' marked pixels denotes the candidate pixels for the second region. The prediction methodology must be very accurate so that very less prediction error occurs. Since the errors need to be expanded so large errors will lead to more distortion. Due to this fact a hybrid predictor is used in this paper which utilizes the variance of the prediction context to predict the pixel value.

The hybrid predictor as given in (2)–(7) makes use of three different predictors on the basis of variance of the prediction context. If variance is less than 10, rhombus predictor is used. A modified version of the Checkerboard Predictor used in [30] is utilized if the variance lies within the range [10 20]. Finally, the middle values of sorted prediction context are used for prediction if variance is greater than 20.

The prediction context for the first region will consist of pixels of the second region and vice versa. To predict the value of pixel $f_{i,j}$, the prediction context, 'c', will be:

$$c = [f_{i-1,j},\ f_{i,j+1},\ f_{i+1,j},\ f_{i,j-1}] \tag{2}$$

$$var(c) = \frac{\sum_{k=1}^{4}(c_k - \bar{c})^2}{3} \tag{3}$$

$$f'_{i,j} = \begin{cases} p, & var(c) \le 10 \\ q, & 10 < var(c) \le 20 \\ r, & \text{elsewhere} \end{cases} \tag{4}$$

$$p = \frac{c(1) + c(2) + c(3) + c(4)}{4} \tag{5}$$

$$q = \begin{cases} \frac{c(2)+c(4)}{2}, & ||c(1) - c(3)|| > ||c(4) - c(2)|| \\ \frac{c(1)+c(3)}{2}, & ||c(1) - c(3)|| < ||c(4) - c(2)|| \\ \frac{c(1)+c(2)+c(3)+c(4)}{4}, & \text{otherwise} \end{cases} \tag{6}$$

$$r = \frac{cs(3) + cs(4)}{2} \tag{7}$$

where 'var(c)' denotes the variance of the prediction context and 'cs' is the prediction context c sorted in the ascending order.

The prediction values obtained in this step can now be used as to calculate the prediction errors, i.e. the difference between the original pixel value and the predicted pixel value.

The prediction errors thus obtained are sorted in ascending order according to the variance of their prediction context. In [14], it has been shown that the performance of the reversible watermarking algorithm in sorted case is better than the unsorted case. For the small variance, prediction errors are also small. Thus the algorithm based on sorted prediction errors exhibits less distortion and results in better quality watermarked image. This can be supported by Fig. 11 also.

3.1.3 Histogram range calculation

Once all the prediction errors are obtained, a histogram of these errors is devised. Instead of using the histogram of difference between the adjacent pixels, prediction error histogram has been used as it is more sharply distributed which leads to lesser distortion. From the histogram, two thresholds- negative threshold t_n and positive threshold t_p are calculated. Only those errors will be expanded which are covered within this range. These thresholds are calculated according to the size of payload. The smaller magnitude prediction errors are considered for the embedding purpose because large prediction errors when expanded will cause large distortion in the watermarked image so initially both t_n and t_p are taken as zero. If the magnitude of zero bin is less than the payload size, the negative threshold is decreased by 1. Again the size of payload will be compared with the total magnitude of histogram bins t_n and t_p. If the size of payload is still more, the positive threshold will be increased by 1. This cycle will be repeated until the total payload size is covered.

3.1.4 Generation of flowmap

Due to the embedding algorithm some pixels might suffer from "data overflow" and "data underflow" problem. Data overflow is the situation when the watermarked pixel exceeds the maximum value of image, i.e. 255 for an 8-bit image and data underflow occurs when this value becomes less than zero [10]. If such pixels are left without embedding, then the decoder would struggle while recovering the image as it would not be able to identify these particular locations. So the information of such pixel locations must be available to the decoder. Keeping in view such problem, a flowmap is created which contains the record of all the potential candidate pixels that might suffer from this problem. These pixels are left unchanged. In this work, the method for flowmap creation is inspired by the method presented in [14].

This methodology works by expanding the prediction error twice. For expansion a test bit is used which is '1' for positive prediction errors and '0' for negative prediction errors. According to the result of expansion these are divided into three categories:

- Category A: It provides information for those prediction errors which experience overflow or underflow just after the first expansion. So these will not be used for expansion. A '0' will be placed in the flowmap to represent these errors.
- Category B: It is used to determine about those prediction errors which experience overflow or underflow after second expansion not the first one. These errors will be expanded for embedding and a '1' will be placed in the flowmap to represent them.

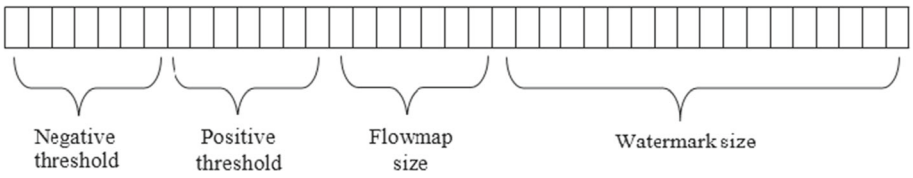


Fig. 4 Structure of payload header

- Category C: If the prediction error does not experience any overflow or underflow even after the second expansion then there is no need to provide any information about them in the flowmap. These errors will be used for watermark embedding.

The flowmap obtained this way does not contain any information about the category C prediction errors. Also the size of category C prediction errors is much greater than the size of category A and category B prediction errors which leads to the reduction in flowmap size considerably as compared to the other existing techniques.

During extraction the prediction errors are expanded only once. If the prediction error does not suffer from overflow or underflow, then that error can be expanded. If it then flowmap is checked and bit '0' of flowmap implies no embedding while bit '1' implies embedding is done.

3.1.5 Embedding of payload header

As discussed earlier the parameters such as t_n, t_p, flowmap size and watermark size are prerequisite for the decoder, so an array is created by concatenating all of these. This array, known as the payload header as shown in Fig. 4, is to be embedded in the image using LSB replacement. For the first region, the LSBs of the pixels of first row of watermarked image are used and for the second region, the LSBs of the pixels of last row of the watermarked image are used. Since we have to extract both the watermark and the original image faithfully at the receiver side, so the LSBs of such pixels are first saved before the actual replacement. After that each bit of the payload header will be replaced with the LSB of these pixels of the image. The final watermark, known as payload, will consist of LSBs of the used pixels, flowmap and the original watermark.

3.1.6 Watermark embedding

The embedding is done using HS and the pairwise expansion of sorted prediction errors as it reduces the distortion and overlapping problem as compared to the other techniques. This further leads to the enhanced embedding capacity of the algorithm.The flowchart shown in Fig. 5 illustrates the steps to obtain the watermarked image. The flowchart represents the embedding methodology only for first region. The same steps are required to be followed with second region to get the final watermarked image. For embedding, two different sets of pixels, set A and set B, are created from the sorted prediction errors. Set A encompasses those prediction errors which occurs within the range of positive and negative threshold and set B incorporates those prediction errors which lie outside this range. This formation of sets is continued until the size of the payload is not achieved. This is given by (8):

$$e_{i,j} = \begin{cases} A, & t_n \leq e_{i,j} \leq t_p \\ B, & \text{otherwise} \end{cases} \tag{8}$$

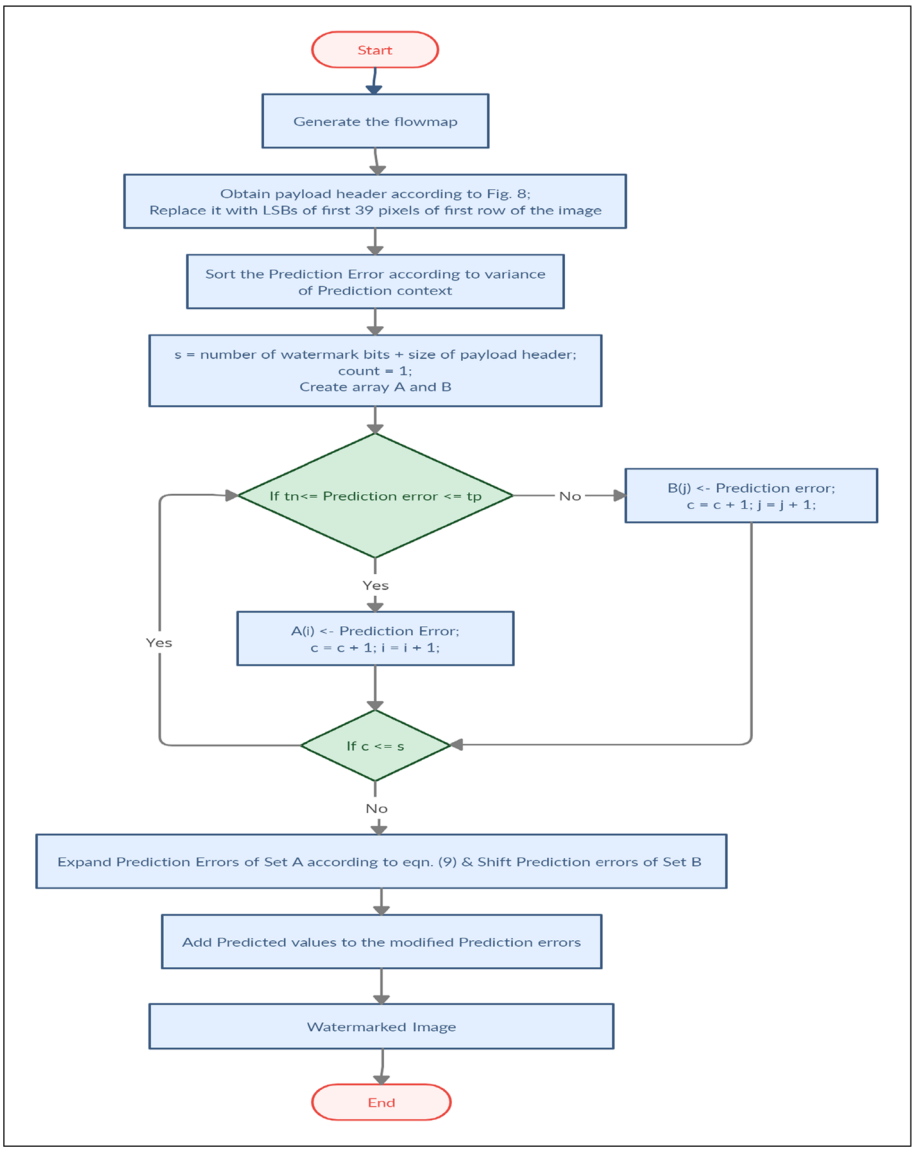


Fig. 5 Flowchart to obtain the watermarked image.It represents the embedding methodology only for the first region. The same steps are required to be followed with the second region also to get the final watermarked image

The payload embedding methodology exploits the pairwise PEE technique. This paper exploits the correlation between the pixels corresponding to the adjacent sorted prediction errors. Since the prediction errors are sorted on the basis of variance, the expansion of adjacent sorted prediction errors will lead to less distortion which in turn will provide high quality watermarked image. The errors of Set A are used to form the pairs and hence for

embedding the watermark. The combination of '00', '01' and '10' will be used to watermark the pair. The combination '11' is not used to reduce the distortion.

The expansion of prediction errors is done according to the (9):

$$e'_{i,j} = \begin{cases} 2e_{i,j} + b, & t_n \leq e_{i,j} \leq t_p \\ e_{i,j} + t_p + 1, & t_p \geq 0 \text{ and } e_{i,j} > t_p \\ e_{i,j} + t_n, & t_n < 0 \text{ and } e_{i,j} < t_n \end{cases} \tag{9}$$

where '$e_{i,j}$' is the prediction error of pixel '$f_{i,j}$', '$e'_{i,j}$' is the expanded prediction error and 'b' is the payload bit that is to be embedded. The value of 'b' can be either '0' or '1'. If it is '1' then it will be '0' for the next pixel of the pair as pair of two pixels is formed and combination of '11' is not allowed to use. The errors lying outside the range, that is errors of Set B, are shifted to prevent the overlapping. From (9), it can be concluded that the expansion of larger prediction errors can cause higher distortion. That's why smaller histogram bins are considered for the expansion of prediction errors. Due to this fact, higher PSNR is possible to achieve even for higher embedding capacity.

The watermarked pixel will be obtained by adding the predicted pixel and the expanded prediction error as given by (10):

$$w_{i,j} = f'_{i,j} + e'_{i,j} \tag{10}$$

where '$w_{i,j}$' represents the watermarked pixel.

The above procedure is repeated for the second region also. The watermarked image obtained after the first region will be considered as the cover image for the second region and the final watermarked image can be obtained using the same procedure.

3.2 Extraction of watermark and cover image

The extraction procedure is the reversal of the embedding procedure as shown in Fig. 6. It considers the region two first followed by region one for extraction. The image recovered from the second region will be considered as watermarked image for the first region. The following sub-sections elaborates the procedure of the watermark extraction method.

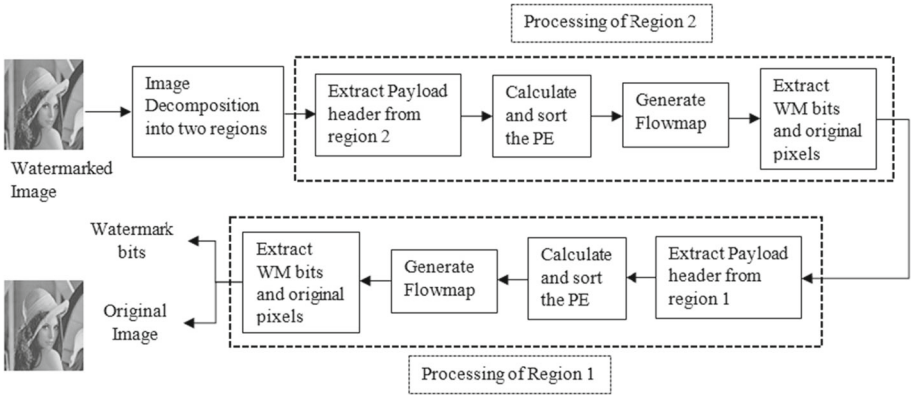


Fig. 6 Procedure to extract watermark and cover image

3.2.1 Image decomposition

The decoding approach also starts with the decomposition of the image into two non-overlapping regions as it has been done during the embedding stage. The Fig. 3 shows the two regions of the image. The purpose also remains same that is to prevent the overlapping of prediction context.

3.2.2 Extraction of payload header

The payload header is extracted from the LSBs of the pixels of the first row of the watermarked image for the region one and LSBs of the pixels of the last row of the watermarked image for the region two. In decoding stage, the positive threshold is $2t_p + 1$ and negative threshold is $2t_n$. The prediction errors lying between these two thresholds will be considered for the extraction of watermark.

3.2.3 Prediction methodology

This step targets to predict the values of the candidate pixels. To realize the distortionless extraction, the same hybrid predictor is utilized as represented by the (2)–(7) and subsequently the prediction errors are calculated. The prediction errors thus obtained are sorted according to the variance of their prediction context.

3.2.4 Watermark extraction and recovery of cover image

The extraction of payload also starts with the formation of Set A and Set B. This time Set A contains the errors lying within the range of $2t_p+1$ and $2t_n$. The errors outside this range constitutes Set B. This formation continues until the payload size is achieved. The errors of Set A are used for payload extraction while those present in Set B are shifted to retrieve the original prediction errors. The extraction is done according to the (11).

$$b = e'_{i,j} \bmod 2, \text{ if } e'_{i,j} \epsilon [2t_n, 2t_p + 1] \tag{11}$$

where, 'b' is the extracted bit and '$e'_{i,j}$' is the prediction error calculated at the extraction stage.

After extracting all the bits, the prediction errors are restored to their original values using the (12):

$$e_{i,j} = \begin{cases} \lfloor e'_{i,j}/2 \rfloor, & 2t_n \le e'_{i,j} \le 2t_p + 1 \\ e'_{i,j} - t_p - 1, & t_p \ge 0 \text{ and } e'_{i,j} > 2t_p + 1 \\ e'_{i,j} - t_n, & t_n < 0 \text{ and } e'_{i,j} < 2t_n \end{cases} \tag{12}$$

The bits retrieved using the (10) are replaced with the LSBs of the pixels of first row for the first region and pixels of the last row of the image for the second region. This restores the original pixels of the respective row.

The original image is extracted by adding the restored prediction errors and the predicted values. The steps discussed so far are repeated for the first region also. The image extracted after the processing of region two will be considered as watermarked image for the region one.

4 Results and discussions

4.1 Experimental results

This section investigates the performance of the proposed reversible watermarking algorithm. The parameter Peak Signal to Noise Ratio (PSNR) is used to check the visual quality of the watermarked image. A PSNR of high value is always desirable. It is given by (14):

$$MSE = \frac{\sum_{p=0}^{M-1} \sum_{q=0}^{N-1} (W_{p,q} - X_{p,q})^2}{M * N} \tag{13}$$

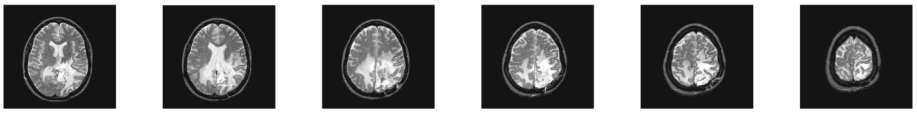
$$PSNR = \log_{10} \left(\frac{Max_X^2}{MSE} \right) \tag{14}$$

where, 'MSE' is the Mean Square Error between the original image and the watermarked image. 'X' is the original image. 'W' is the watermarked image. M*N gives the size of the image. 'Max_X' is the maximum possible value of the image X which is 255 for an 8-bit grayscale image.

Since the maximum MSE can be 1 so the corresponding PSNR will be approximately equal to 48. Due to this reason the embedding capacity is not increased beyond the value that results in PSNR below 48.

The tool used for the implementation of the algorithm is MATLAB R2017a. The proposed algorithm has been executed over 66 images taken from different datasets, namely USC-SIPI dataset, CVG-UGR database (Figs. 7 and 8), hlevkin database and Kodak Image dataset. These test images are of diverse nature, i.e., some of them are standard test images, some are used for military applications and some are used for medical purposes. The performance of the algorithm against these images can be analysed using Fig. 9. Figure 10 shows how visual quality and embedding capacity vary with threshold for Lena image . The effect of sorting the prediction errors on visual quality can be understood using Fig. 11. Figures 12, 13 and 14 compares the performance of the suggested algorithm with the existing approaches against the frequently used standard test images, as shown within the plot. The standard grayscale test images used to evaluate the performance of the proposed algorithm are Lena, Mandrill, Airplane and Lake as shown in Fig. 12. Since the reversible watermarking finds its application in the biomedical, medical, aerial and military images, hence some of these kinds of images are also used to measure the algorithm's performance. The images used to test the performance against biomedical and medical images are Algae, Fingerprint, Heart and X-ray as shown in Fig. 13. To assess the performance of the proposed algorithm in the aerial and military applications, the images used are aerial, military airplane, military APC and military car and APCs. These images are shown in Fig. 14. All these images are grayscale with size of 512*512 pixels. Furthermore, to compare the performance of the proposed algorithm with the existing state-of-the- art techniques, it is executed on colour

Fig. 7 Standard grayscale test images of size 512*512 pixels taken from CVG-UGR database: (a) Barbara, (b) Town, (c) Objects, (d) Monument, (e) Bird, and (f) Sailboat



(a) (b) (c) (d) (e) (f)

Fig. 8 Standard grayscale medical MR test images of size 256*256 pixels taken from the CVG-UGR database: (a) MR1, (b) MR2, (c) MR3, (d) MR4, (e) MR5, and (f) MR6

images of Kodak Image dataset which are first transformed to grayscale images before the further processing. The payload size of 30,000 bits have been taken to compare the results with the existing techniques as represented in the Table 2.

4.2 Effect of sorting the prediction errors

The proposed algorithm sorted the prediction error on the basis of variance of prediction context to reduce distortion at smaller payloads. The smaller value of variance implies that the diversity within the neighbourhood of a pixel is not high. This facilitates the predicted value to be closer to the original one and hence the smaller prediction error. Moreover, embedding in smaller prediction error leads to high quality watermarked image. To verify this theory, the algorithm was tested both ways- with and without sorting. The experiments conducted on some of the standard test images have proven the fact and the same has been illustrated in Fig. 11. The graph shows that there exists a considerable difference between the PSNR obtained in the two cases specifically for the smaller payload. Therefore, sorted case has been considered further for the evaluation of the performance of algorithm.

4.3 Visual quality vs embedding capacity

This section presents the results obtained after the implementation of the proposed algorithm. The x-axis of the graph denotes the Embedding Capacity of the algorithm while the y-axis represents the PSNR in terms of dB which depicts the visual quality of the watermarked image. The higher the value of PSNR, the better is the visual quality of the watermarked image. As discussed above, the maximum embedding capacity is that number of bits whose corresponding PSNR is 48.

The algorithm has been tested against a number of images of different datasets. Figure 9 depicts the PSNR obtained for different embedding capacity for these images. The size of images are 512 * 512 pixels except for MR images of CVG-UGR database. The size of MR images is 256 * 256 pixels. From Fig. 9, it can be concluded that the proposed algorithm is capable of achieving high embedding capacity as the maximum embedding capacity lies within the range of 15% to 76% approximately. Most of the images have shown higher embedding capacity.

The performance of the algorithm is compared with the algorithms presented by Ou et al. [27], Dragoi et al. [8], Zhang et al. [40], Kouhi [17], M. Ishtiaq et al. [14], Li et al. [20] and Naskar et al. [24] and results are presented in Fig. 12. Figure 12(a) represents the results for Lena image which shows that higher PSNR is achieved and the embedding capacity is considerable high than the existing algorithms. Figure 12(b) illustrates the results for Mandrill image. At smaller payload the algorithm provides a slightly smaller PSNR but it increases with payload and achieves high embedding capacity. Figure 12(c) depicts the results for Airplane image which outperforms the existing techniques both in terms of

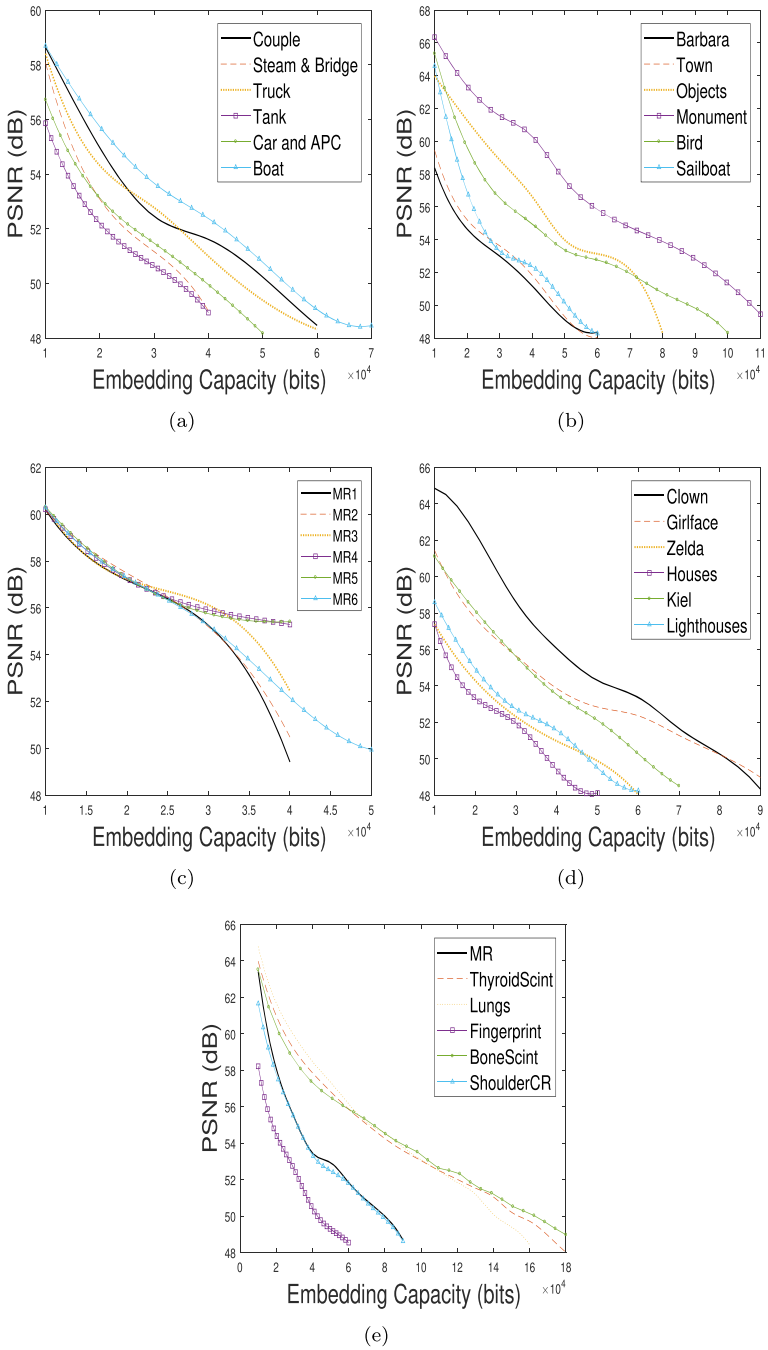


Fig. 9 Embedding Capacity vs PSNR for (a) test images of USC-SIPI dataset, (b) test images of CVG-UGR database, (c) MR images of CVG-UGR database, (d) test images of hlevkin database, and (e) medical images of hlevkin database

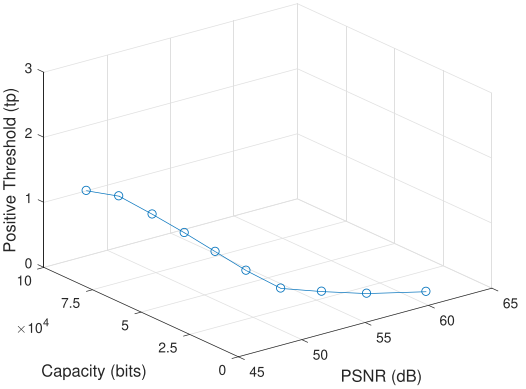


Fig. 10 Variation of Visual quality and embedding capacity with threshold for Lena Image

distortion and embedding capacity. Figure 12(d) represents the results for Lake image. It shows that initially PSNR is same but it increases considerably at higher payloads and the embedding capacity is also very high than the other algorithms.

The algorithm has shown appreciable results when applied to the standard grayscale test images. But the reliability of the algorithm must be tested from the application point of view also. One of the most important applications of reversible watermarking is in the field of medical and biomedical images where the integrity of original image is equally important as

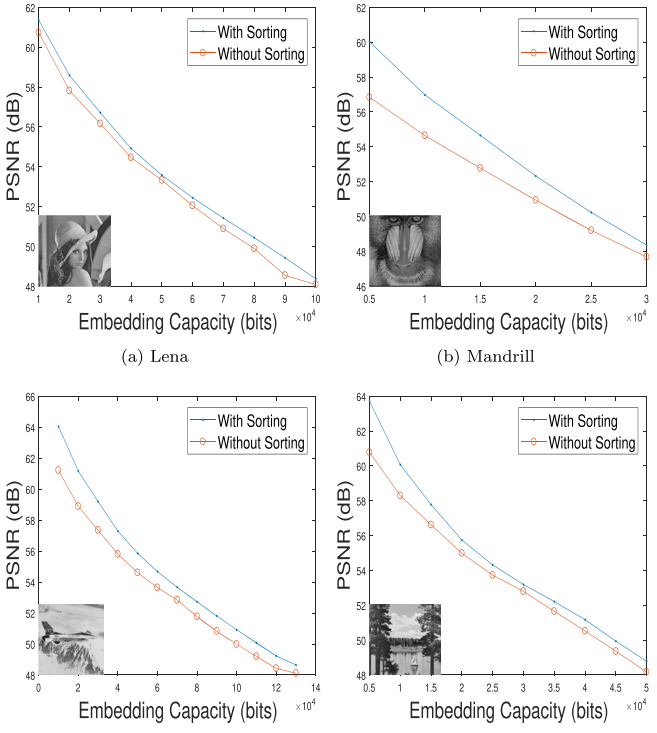


Fig. 11 Effect of watermark embedding with and without the sorting of prediction errors tested on standard grayscale test images

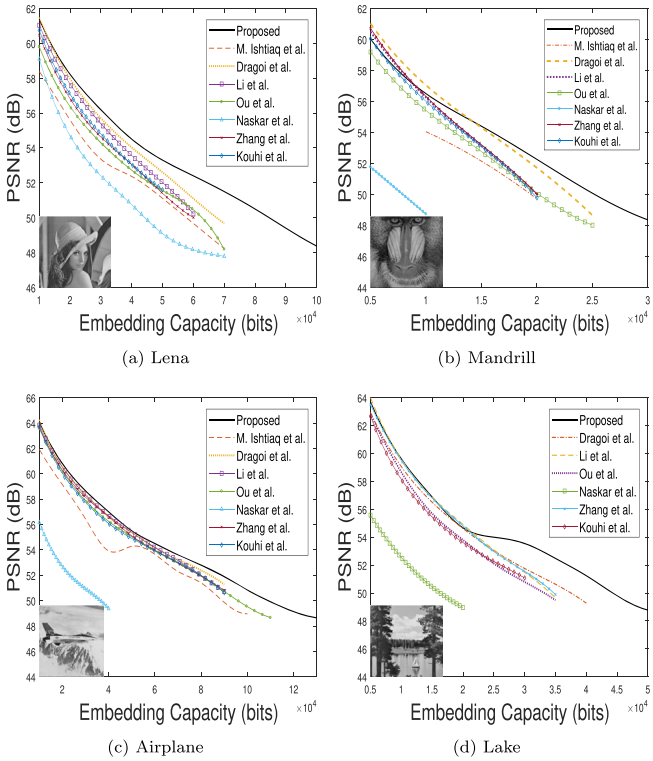


Fig. 12 Performance of proposed algorithm tested on standard grayscale test images and compared with the algorithm presented by Ou et al. [27], Dragoi et al. [8], Zhang et al. [40], Kouhi [17], M. Ishtiaq et al. [14], Li et al. [20] and Naskar et al. [24]

that of the watermark. For this the various medical and biomedical images have been taken to observe the performance of the algorithm.

The results obtained in the case of biomedical and medical images have been plotted in Fig. 13 and compared with Naskar et al. scheme [24]. The proposed scheme outperforms the algorithm present in [24] in terms of both distortion and embedding capacity except in case of algae where for smaller payload size scheme [24] is better. When the payload size increases the proposed algorithm performs better. Both the PSNR and the embedding capacity is significantly high. This is because of the sorting of the prediction errors using their variance and embedding the watermark in the lower magnitude histogram bins.

Another important application of the reversible watermarking is found to be in the aerial and military images where the exact retrieval of both the watermark and the original image is given utmost importance due to the security point of view. Thus the algorithm's performance is evaluated against various aerial and military images and the results have been shown in Fig. 14. The results have shown the superiority of the proposed algorithm over the algorithm presented by Naskar et al. [24].

Table 2 depicts the performance of the proposed approach when evaluated against the colour images of the Kodak dataset. The results are compared with the methodologies proposed by Xiong [37], Lee et al. [18], Hong et al. [11], Fallahpour [9], Yang et al. [38], Tai et al. [33], Li et al. [19], Yanga et al. [39], Hu et al. [13] and Wahed et al. [36] for the

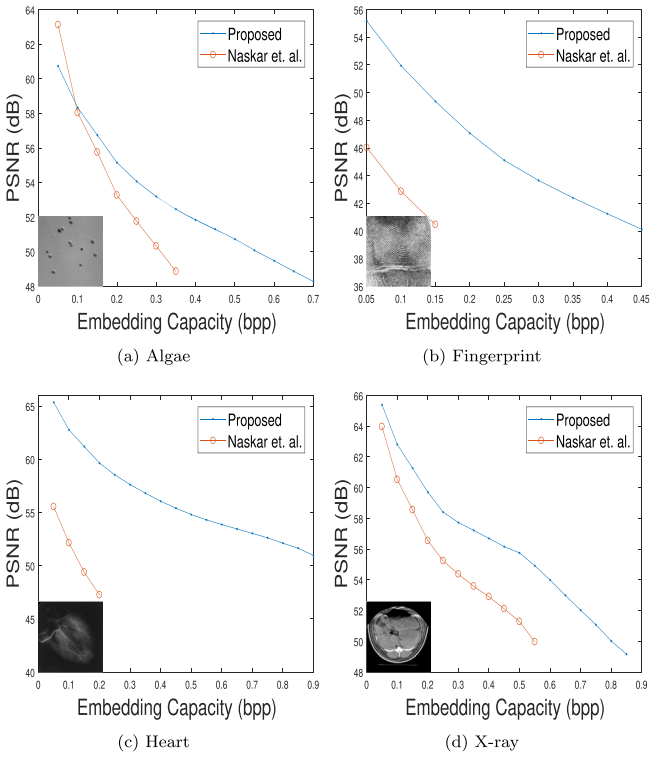


Fig. 13 Performance of proposed algorithm tested on Biomedical and Medical images and compared with the algorithm presented by Naskar et al. [24]

same embedding capacity of 30,000 bits. The average PSNR for such embedding capacity is 57.3232 as obtained from Table 2. This is greater than the existing approaches. So it can be interpreted that the proposed algorithm outperforms the existing techniques for colour images also.

4.4 Complexity of the algorithm

The proposed algorithm is less computationally complex if compared to the approach presented by Dragoi et al. [8]. This can be substantiated by the fact that the proposed algorithm forms pair between the adjacent sorted prediction errors to reduce the distortion while embedding the watermark bits. Whereas approach proposed in [8] forms the pairing of the candidate pixel with one of its neighbouring pixels according to the local complexity. This requires more processing to compare the neighbouring pixels as compared to the proposed algorithm. Moreover, the proposed methodology while generating the flowmap doesn't incorporate the locations where watermark bit is embedded. Had these positions have also been included then some compression technique must be required to reduce the flowmap size which is not the case with the proposed algorithm. This suggests that the algorithm is computationally effective also.

- Time Complexity

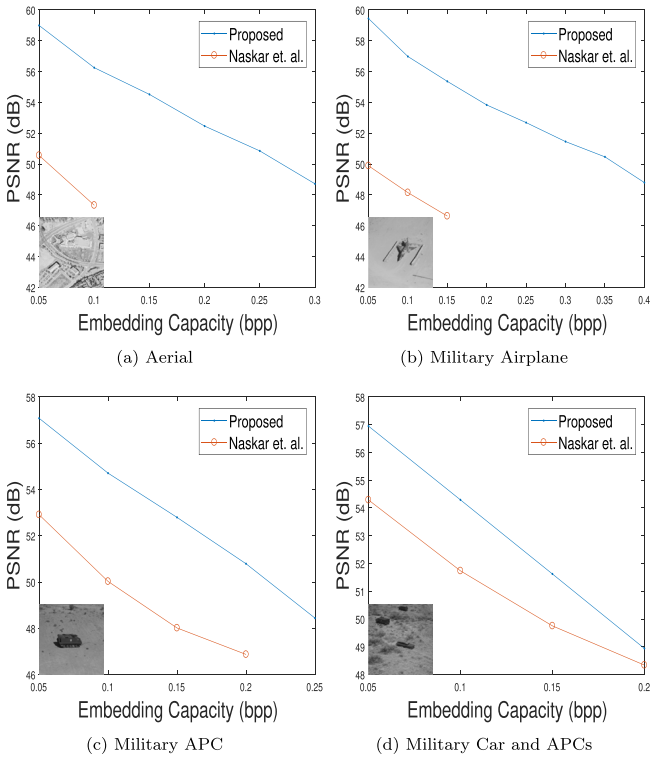


Fig. 14 Performance of proposed algorithm tested on Aerial and Military images and compared with the algorithm presented by Naskar et al. [24]

1. The initialization of image needs O($I_i m$ (mxn)) time where $I_i m$ represents original image, m represents the number of rows and n represents the number of columns in each image.
2. The time taken by processing region one is O(P_1 (mxn)) time where P_1 represents region 1 of the image.
3. The time taken by processing region two is O(P_2 (mxn)) time where P_2 represents region 1 of the image.
4. The total time taken by the watermark embedding and watermark extraction procedure is 2*(O(P_1 (mxn)) + O(P_2 (mxn))). For extraction purpose also the two regions of image have been considered.

- Space Complexity

  The original image and the watermarked image are required during the watermark embedding stage and watermark extraction stage, respectively. So the total space complexity is O($I_i m$ (mxn)) + O($W_i m$ (mxn)) where $W_i m$ represents the watermarked image.
- Execution Time

  The algortihm is implemented on the Intel Core i3 processor and 4GB RAM using MATLAB R2017a. The execution time taken in the processing of the standard grayscale test images is depicted in the Fig. 15.

Table 2 Performance of proposed algorithm tested on Kodak Image dataset for embedding capacity of 30,000 bits and compared with the algorithm presented by [9, 11, 13, 18, 19, 33, 36–39]

Images	Proposed	Xiong [37]	Lee and Chen [18]	Hong et al. [11]	Fallahpour [9]	Yang and Tsai [38]	Tai et al. [33]	Li et al. [19]	Yanga et al. [39]	Hu and Li [13]	Wahed and Nyeem [36]
Kodim01	52.6414	52.2959	–	50.4266	51.2073	51.6321	48.5875	50.3778	47.3665	40.5227	53.3899
Kodim02	57.6832	54.3170	50.8657	52.3653	53.1956	54.1495	50.8720	52.7678	51.2495	45.2383	56.7655
Kodim03	59.7805	57.2181	54.1298	55.1091	56.6645	57.5689	54.6394	56.8184	48.1877	43.4133	54.8446
Kodim04	57.1624	54.9052	51.0863	52.6107	53.5588	54.2580	50.3465	52.3079	48.6195	40.9479	52.2864
Kodim05	55.9337	54.3859	49.9140	52.0451	51.9686	53.4338	49.1237	51.7606	45.6954	39.9903	55.5589
Kodim06	59.5788	58.2658	55.9930	56.2324	58.5470	57.4635	56.7478	58.2100	43.9294	38.1768	52.5960
Kodim07	58.8416	56.8830	53.2429	54.7846	55.2287	57.2452	53.4903	55.9154	50.7966	45.4427	54.7184
Kodim08	52.2394	51.9491	–	49.9693	51.2096	49.9143	48.3205	49.3666	46.7856	37.9303	52.9210
Kodim09	55.7015	55.3468	51.8038	53.6817	54.3293	55.0885	52.1903	54.9489	52.0286	50.3558	54.7022
Kodim10	56.0477	55.0519	51.4149	53.4783	53.9502	54.4459	51.5444	53.9102	48.9295	43.2751	55.7438
Kodim11	58.2231	57.0155	52.4497	54.3926	55.0963	55.8792	53.5448	56.7322	48.6685	45.0744	57.0435
Kodim12	58.1673	56.7206	51.5439	53.8004	53.9889	55.5933	52.3606	55.7170	47.0276	39.7519	52.1662
Kodim13	51.3574	50.7106	–	48.3305	–	49.5127	–	48.5430	46.6910	39.1366	52.4139
Kodim14	54.4162	53.7185	49.1863	51.1071	51.5217	53.0315	49.2509	51.3486	49.2799	42.0084	55.9105
Kodim15	60.5123	54.6280	51.7366	53.1060	54.3064	54.8403	52.0906	53.7393	48.9915	41.5032	54.3856
Kodim16	58.6685	57.5072	54.1787	55.4802	56.2210	57.7042	55.1016	57.4098	50.4266	47.3475	56.3419
Kodim17	56.6431	56.0587	52.1193	53.9419	54.1182	55.5784	52.2809	55.3358	46.0879	43.0160	56.7109
Kodim18	53.0576	52.7248	48.9868	50.5623	–	51.7264	–	50.4577	50.9331	44.2985	57.4373
Kodim19	56.1943	55.9403	52.5382	54.2026	54.2740	55.7769	52.3299	55.0635	52.6711	55.0883	55.7460
Kodim20	71.9422	55.1350	57.7668	55.4947	57.9826	52.9519	58.6097	52.3295	43.9391	34.6423	49.3229
Kodim21	56.0392	55.5335	52.2829	54.0467	54.4882	55.6224	52.6592	55.3512	51.8352	58.3449	56.7221
Kodim22	55.5372	53.8981	49.7114	51.5627	52.2460	52.6712	49.2655	52.2160	50.3458	47.4917	52.0387
Kodim23	58.6353	56.3887	53.4745	54.7149	55.6291	56.5624	53.0857	55.5898	57.1935	54.5841	55.2145
Kodim24	60.7538	54.6610	51.4304	52.1461	54.0749	52.8682	51.7569	52.2730	45.3868	38.2884	50.0820

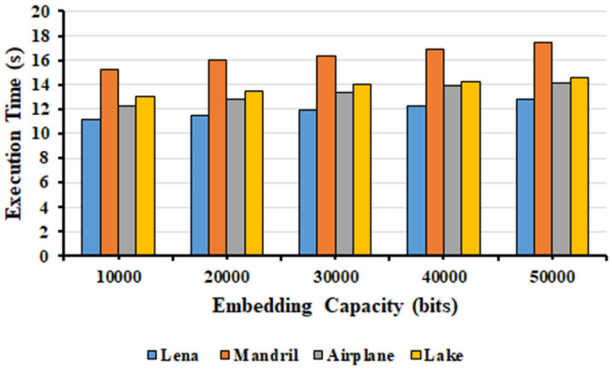


Fig. 15 Variation of execution time with embedding capacity for test images: Lena, Mandril, Airplane and Lake

5 Conclusion

In this paper, a fragile reversible watermarking algorithm has been proposed and implemented successfully. The algorithm is based on pairwise prediction error expansion and histogram shifting which claims to exhibit less distortion and high embedding capacity. The results obtained indicate the lesser distortion even at smaller payloads. The algorithm is able to extract the watermark and the host image without any distortion. The results further portray that the proposed algorithm outperforms the existing ones thus making it more suitable for medical, military, aerial images.

The proposed algorithm can further be made to work efficiently for video as well in future.

Data Availability The authors have taken the images from the following Image Databases for the proposed work:

1. USC-SIPI Image Database available online at: https://sipi.usc.edu/database/
2. CVG-UGR Image Database available online at : https://ccia.ugr.es/cvg/dbimagenes/index.php
3. hlevkin Image Database available online at: http://www.hlevkin.com/hlevkin/06testimages.htm
4. Kodak Image Database available online at: http://r0k.us/graphics/kodak/

Declarations

Competing interests The authors declare that they do not have any competing financial interests nor any personal relationships that could seem o have influenced the work presented in this paper.

References

1. Abdulla AA (2015) Exploiting similarities between secret and cover images for improved embedding efficiency and security in digital steganography. The University of Buckingham, UK
2. Agilandeeswari L, Ganesan K (2016) A robust color video watermarking scheme based on hybrid embedding techniques. Multimed Tools Appl 75:8745–8780. Springer
3. Agilandeeswari L, Ganesan K (2016) An efficient hilbert and integer wavelet transform based video watermarking. J Eng Sci Technol 11:327–345
4. Agilandeeswari L, Ganesan K (2018) RST invariant robust video watermarking algorithm using quaternion curvelet transform. Multimed Tools Appl 77:25431–25474. Springer
5. Ahmaderaghi B, Kurugollu F, Rincon JMD et al (2018) Blind image watermark detection algorithm based on discrete shearlet transform using statistical decision theory. IEEE Trans Comput Imaging 4:46–59

6. Celik MU, Sharma G, Tekalp AM et al (2005) Lossless generalized-LSB data embedding. IEEE Trans Image Process 14:253–266
7. Chang CS, Shen JJ (2017) Features classification forest: a novel development that is adaptable to robust blind watermarking techniques. IEEE Trans Image Process 26:3921–3935
8. Dragoi IC, Coltuc D (2016) Adaptive pairing reversible watermarking. IEEE Trans Image Process 25:2420–2422
9. Fallahpour M (2008) Reversible image data hiding based on gradient adjusted prediction. IEICE Electron Express 5:870–876
10. Feng B, Yu B, Yilin B et al (2019) A reversible watermark with a new overflow solution. IEEE Access 7:28031–28043
11. Hong W, Chen TS, Shiu CW (2009) Reversible data hiding for high quality images using modification of prediction errors. J Syst Softw 82:1833–1842
12. Honsinger CW, Jones PW, Rabbani M et al (2001) Lossless recovery of an original image containing embedded data. U.S. Patent No. 6,278,791
13. Hu J, Li TR (2015) Reversible steganography using extended image interpolation technique. Comput Elect Eng 46:447–455
14. Ishtiaq M, Ali W, Shahzad W et al (2018) Hybrid predictor based four-phase adaptive reversible watermarking. IEEE Access 6:13213–13230
15. Jung KH (2018) A survey of interpolation-based reversible data hiding methods. Multimed Tools Appl 77:7795–7810. Springer
16. Khan A, Siddiqa A, Munib S et al (2014) A recent survey of reversible watermarking techniques. Inform Sci 279:251–272. Elsevier
17. Kouhi A, Sedaaghi MH (2022) Reversible data hiding based on high fidelity prediction scheme for reducing the number of invalid modifications. Inf Sci 589:46–61
18. Lee CF, Chen HL (2012) Adjustable prediction-based reversible data hiding. Digit Signal Process 22:941–953
19. Li YC, Yeh CM, Chang CC (2010) Data hiding based on the similarity between neighboring pixels with reversibility. Digit Signal Process 20:1116–1128
20. Li X, Zhang W, Gui X et al (2016) Efficient reversible data hiding based on multiple histograms modification. IEEE Trans Inf Forensics Sec 10:2016–2027
21. Lin CC, Tai WL, Chang CC (2008) Multilevel reversible data hiding based on histogram modification of difference images. Pattern Recognit 41:3582–3591. Elsevier
22. Mandal PC, Mukherjee I, Paul G, Chatterji BN (2022) Digital image steganography: a literature survey. Inform Sci 609:1451–1488. Elsevier
23. Memon NA, Khan A, Gilani SAM et al (2011) Reversible watermarking method based on adaptive thresholding and companding technique. Int J Comput Math 88:1573–1594
24. Naskar R, Chakraborty RS (2012) Reversible Watermarking utilising weighted median-based prediction. IET Image Process 6:50–520
25. Ni Z, Shi YQ, Ansari N et al (2006) Reversible data hiding. IEEE Trans Circ Syst Video Technol 16:354–362
26. Ortiz AM, Uribe CF, Beltran RH et al (2019) A survey on reversible watermarking for multimedia content: a robustness review. IEEE Access 7:132662–132681
27. Ou B, Li X, Zhao Y et al (2013) Pairwise prediction-error expansion for efficient reversible data hiding. IEEE Trans Image Process 22:5010–5021
28. Ou B, Li X, Wang J (2016) High-fidelity reversible data hiding based on pixel-value-ordering and pairwise prediction-error expansion. J Vis Commun Image Represent 39:12–23. Elsevier
29. Poonam, Arora SM (2018) A DWT-SVD based robust digital watermarking for digital images. Procedia Comput Sci 132:1441–1448. Elsevier
30. Rad RM, Wong KS, Guo JM (2016) Reversible data hiding by adaptive group modification on histogram of prediction errors. Signal Process 125:315–328. Elsevier
31. Sachnev V, Kim HJ, Nam J et al (2009) Reversible watermarking algorithm using sorting and prediction. IEEE Trans Circ Syst Video Technol 19:989–999
32. Samee R, Riaz MM, Ghafoor A (2018) Adaptive interpolation and segmentation based reversible image watermarking. Multimed Tools Appl 77:26821–26843. Springer
33. Tai WL, Yeh CM, Chang CC (2009) Reversible data hiding based on histogram modification of pixel differences. IEEE Trans Circ Syst Video Technol 19:906–910
34. Thodi DM, Rodriguez JJ (2007) Expansion embedding techniques for reversible watermarking. IEEE Trans Image Process 16:721–30
35. Tian J (2003) Reversible data embedding using a difference expansion. IEEE Trans Circ Syst Video Technol 13:890–896

36. Wahed MA, Nyeem H (2019) High capacity reversible data hiding with interpolation and adaptive embedding. PLOs ONE 14:Art. no e0212093
37. Xiong X (2019) Novel scheme of reversible watermarking with a complementary embedding strategy. IEEE Access 7:136592–136603
38. Yang CH, Tsai MH (2010) Improving histogram-based reversible data hiding by interleaving predictions. IET Image Process 4:223–234
39. Yanga CN, Hsua SC, Kim C (2017) Improving stego image quality in image interpolation based data hiding, vol 50
40. Zhang C, Ou B (2021) Reversible data hiding based on multiple adaptive two-dimensional prediction-error histograms modification. IEEE Trans Circ Syst Video Technol. https://doi.org/10.1109/TCSVT.2021.3125711
41. Zhou J, Au OC (2012) Determining the capacity parameters in PEE-based reversible image watermarking. IEEE Signal Proc Lett 19:287–290

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Journal of Low Temperature Physics (2023) 213:92–106
https://doi.org/10.1007/s10909-023-02990-2

Hydrostatic Pressure Effect on the Thermodynamic Properties of Quantum Wire Under a Crossed Electromagnetic Field

Pranay Khosla[1] · Sakshi Arora[1] · Yash Gupta[1] · Priyanka[1] · Rinku Sharma[1]

Received: 20 April 2023 / Accepted: 18 July 2023 / Published online: 12 August 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
In this paper, a GaAs quantum wire is considered with a semi-parabolic lateral confinement potential in the presence of a crossed electromagnetic field with the inclusion of Rashba SOI. The effects of pressure on thermodynamic properties have been studied. To this end, first, the eigenenergies for the system have been calculated considering Rashba spin effects under electric and magnetic fields. Then a numerical analysis is done to derive the partition function and obtain the system's thermodynamic properties. The results show that at low temperatures, all four thermodynamic properties don't show much effect due to pressure. It is found that entropy increases with increasing hydrostatic pressure, whereas pressure affects mean energy and free energy in a different manner. A peak structure is observed in specific heat as a function of temperature where the peak position depends on hydrostatic pressure.

Keywords Hydrostatic pressure · Quantum wire · Rashba effect · Thermodynamic properties

1 Introduction

Nanostructures have been the focus of many studies in the past decade, owing to their potential applications in micro-electronics, bio-medicine, photonics, quantum computing, and other engineering fields [1–3]. Advancement in several physical methods viz. lithography, molecular beam epitaxy, vapor deposition techniques, etc. [4–6] has made the fabrication of nanostructures feasible. Using these techniques, semiconductors can be grown into low-dimensional structures of desired geometry and size [7, 8], creating structures such as dots, wires, and rings.

✉ Rinku Sharma
rinkusharma@dtu.ac.in

[1] Department of Applied Physics, Delhi Technological University, Delhi 110042, India

Of the nanostructures, quantum wires are promising candidates for low-dimensional conducting wires in electronic, optoelectronic, and quantum computing devices [9]. Quantum wires have electron motion confined in two dimensions for which, various models of confinement potential, viz. parabolic, Rosen–Morse, spherical Gaussian, power-exponential, modified Gaussian, etc., have been proposed [10–13]. The confinement into low dimensions results in the formation of discrete energy levels [14]. At the nanoscale, spin effects also come into consideration. There are two types of Spin Orbital Couplings; Rashba [15], which arises due to structural inversion asymmetry, and Dresselhaus [16], which arises due to bulk inversion asymmetry. Experimental work suggests that the strength of the Rashba effect can be controlled by an external gate voltage [17, 18], thus making it an important part of various studies. The choice of confinement potential, formation of discrete energy levels, and the splitting of energy levels due to spin-orbital interactions affect the Hamiltonian which in turn reflects in the physical properties of the system.

Lately, there have been many studies on the electronic [19, 20], magnetic [21, 22], and optical [23, 24] properties of quantum dots and wires. Due to the possibility of manipulating electron states via external gate voltage, the inclusion of Rashba spin–orbit coupling/interaction (SOI) has been popular in many works [19, 21, 22, 24–28].

Thermodynamic properties are also important to describe the physical state of a system and have been studied extensively. Khordad et al. have analyzed the effect of Rashba SOI, on the thermodynamic properties of quantum wires at low temperatures [26] and the thermodynamics of double ring-shaped quantum dots at both low and high temperatures [29]. Najafi et al. studied mono-layer quantum wire thermodynamics under SOI [28]. A few studies have explored the effects of hydrostatic pressure and impurities on binding energy, ballistic conductance, optical properties, Tsallis entropy, etc., employing an effective mass variation model [19, 23, 30–32]. Khordad, Fizoori, and Sedehi [30] have investigated the simultaneous effects of pressure and temperature on entropy and specific heat of a three-dimensional quantum wire.

We focus on the effects of pressure on the thermodynamic properties of a quantum wire under the influence of Rashba SOI, using a variational mass model. The model is described in Sect. 2 with details about the Hamiltonian of the system. A solution of Schrödinger's equation gives the eigenenergies which are used to derive the partition function and subsequently the thermodynamic properties of the system. In Sect. 3, the results are discussed in light of how pressure affects the physical properties. We conclude in Sect. 4 discussing the effect of pressure in a broader sense with some applications of the study.

2 Theory

The study is formulated for an electron gas confined in the x–y plane. A semi-parabolic lateral confinement potential is applied (Fig. 1a), quantizing the electron motion in the x-direction. The confinement potential is given by $V(x)$. Thus, the quantum wire extends along the y-direction. The Hamiltonian of one electron in the system when subjected to an external magnetic field, $\vec{B} = (0, 0, B)$ including Rashba SOI is given by [24, 25]:

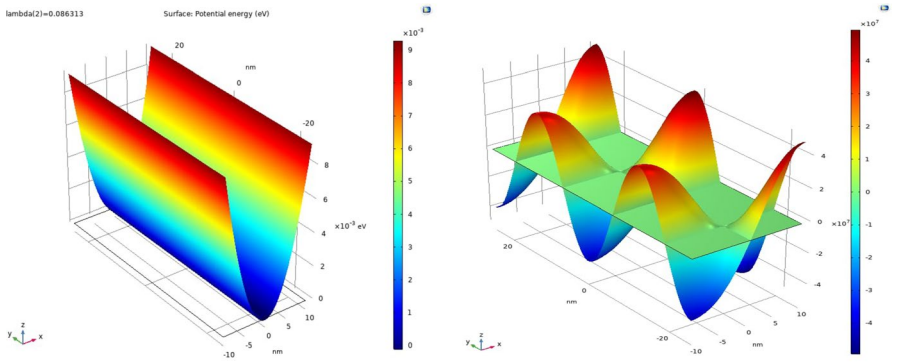


Fig. 1 Surface **a** potential energy (eV) and, **b** normalized wave function of quantum wire

$$H = \frac{1}{2m_{\text{eff}}(P)}(p + eA)^2 + V(x) + \frac{1}{2}g\mu_B\vec{\sigma}.\vec{B} + H_{\text{Rashba}} \tag{1}$$

$$\begin{aligned} V(x) &= \frac{1}{2}m_{\text{eff}}(P)\omega_0^2 x^2, \quad \text{for} \quad 0 \le x \le L/2 \\ &= \infty, \quad \text{for} \quad L/2 \le x \le L \end{aligned} \tag{2}$$

here, $\frac{1}{2}g\mu_B\vec{\sigma}.\vec{B}$ is the Zeeman coupling under magnetic field, p is the momentum, e is the electronic charge ($1.6 \times 10^{-11}C$) and A is the vector potential for the magnetic field, given by $A = eB\hat{xj}$ in Landau gauge. ω_0 is the frequency of confinement, g and μ_B are Lande's g-factor for the electron and the Bohr magneton respectively, and σ represents the Pauli spin matrices (σ_x;σ_y;σ_z). In Eq. (2), $m_{\text{eff}}(P)$ represents the pressure dependent effective mass.

In this paper, an effective mass approximation is used where the effective mass is dependent on hydrostatic pressure. For GaAs, the following formulation gives the hydrostatic pressure dependency of effective mass [31]:

$$m_{\text{eff}}(P) = \left[1 + E_P\left(\frac{2}{E_g(P,0)} + \frac{1}{E_g(P,0) + 0.341}\right)\right]^{-1} m_e \tag{3}$$

where m_e is the free mass of an electron (given by, $m_e = 9.11 \times 10^{-31}kg$), $E_P = 7.51eV$, $E_g(P,T)$ is the energy gap for GaAs which depends on pressure and temperature (here T=0 K) and is given by:

$$E_g(P,0) = {E_g}^0 + \alpha P + \beta P^2 \tag{4}$$

here ${E_g}^0 = 1.519eV, \alpha = 1.26 \times 10^{-2}eV/kbar, \beta = 3.771.26 \times 10^{-2}eV/kbar^2$

In this study, an electric field of strength $\vec{E} = (E_x, 0, 0)$ is applied, which transforms the Hamiltonian (Eq. 1) into [25]:

$$H = H_0 + H_{\text{Rashba}} \tag{5}$$

$$H_0 = \frac{1}{2m_{\text{eff}}(P)}({p_x}^2 + (p_y + eB_z x)^2) + \frac{1}{2}m_{\text{eff}}(P)\omega^2 x^2 + eE_x x + \frac{1}{2}g\mu_B B\sigma_z \tag{6}$$

$$H_{\text{Rashba}} = \frac{\alpha}{\hbar}(\sigma_x(p_y + eB_z x) - \sigma_y p_x) \tag{7}$$

where the effective cyclotron frequency, ω is given by $\omega = \sqrt{{\omega_0}^2 + {\omega_c}^2}$ and, $\omega_c = \frac{eB}{m_{\text{eff}}}$ is the cyclotron frequency.

Due to the inclusion of Pauli spin matrices, the Hamiltonian becomes a 2×2 matrix. From Eq. (6), the Hamiltonian is found to be invariant along the wire, i.e. translationally invariant. Thus, the wave function (Fig. 1b) can be written as a plane wave function in the longitudinal direction, and a spinor for the transversal coordinate, x [33].

$$\Psi(x, y) = exp(ik_y y)\phi(x) \tag{8}$$

where $exp(ik_y y)$ is the plane wave function, $\phi(x)$ is a spinor and k_y is the electron wave vector or the wave number, in the y-direction [24–26]. The eigenvalues for H_0 can be found by solving Schrödinger's equation:

$$H_0\Psi_{n\sigma}(x) = E_{n\sigma}\Psi_{n\sigma}(x) \tag{9}$$

The corresponding eigenenergies are given as

$$E_{n\sigma} = \hbar\omega\left(n + \frac{3}{2}\right) - \frac{e^2E^2}{2m_{\text{eff}}(P)\omega^2} + \frac{\omega_0^2\hbar^2k_y^2}{2m_{\text{eff}}(P)\omega^2} - \frac{e^2EB\hbar k_y}{m_{\text{eff}}^2(P)\omega^2} + \frac{1}{2}g\mu_B\sigma_z B_z \tag{10}$$

$\Psi_{n\sigma}$ are the corresponding wave functions for, $n = 0, 1, 2...$ and $\sigma = \pm 1$.

The eigenvalue equation for the Hamiltonian H can be rewritten by writing $\phi(x)$ as, $\phi(x) = \sum_{n\sigma} c_{n\sigma}\Psi_{n\sigma}(x)$ as

$$\sum_{n\sigma} c_{n\sigma}(E_{n\sigma} - E)\Psi_{n\sigma}(x) + \sum_{n\sigma} c_{n\sigma}H_{\text{Rashba}}\Psi_{n\sigma}(x) = 0 \tag{11}$$

Upon applying the orthogonality condition we get,

$$(E_{n\sigma} - E)c_{n\sigma} + \sum_{n\prime\sigma\prime} c_{n\prime\sigma\prime}\langle\Psi_{n\sigma}|H_{\text{Rashba}}|\Psi_{n\prime\sigma\prime}\rangle = 0 \tag{12}$$

Figure 2 shows the splitting of the energy levels due to Rashba SOI. Here, the solid curves represent the wave function, dashed lines represent eigenenergies corresponding to the wave function of the same color. The black parabolic curve is for the potential. It can be observed that curves 'a' and 'b' have a similar shape, the only difference being in their values. Similarly curves 'c' and 'd', etc., also have similar shapes but different values. This implies that these pairs result from the splitting of eigenenergies at individual energy levels, the extent of the splitting being ~ 0.13 eV.

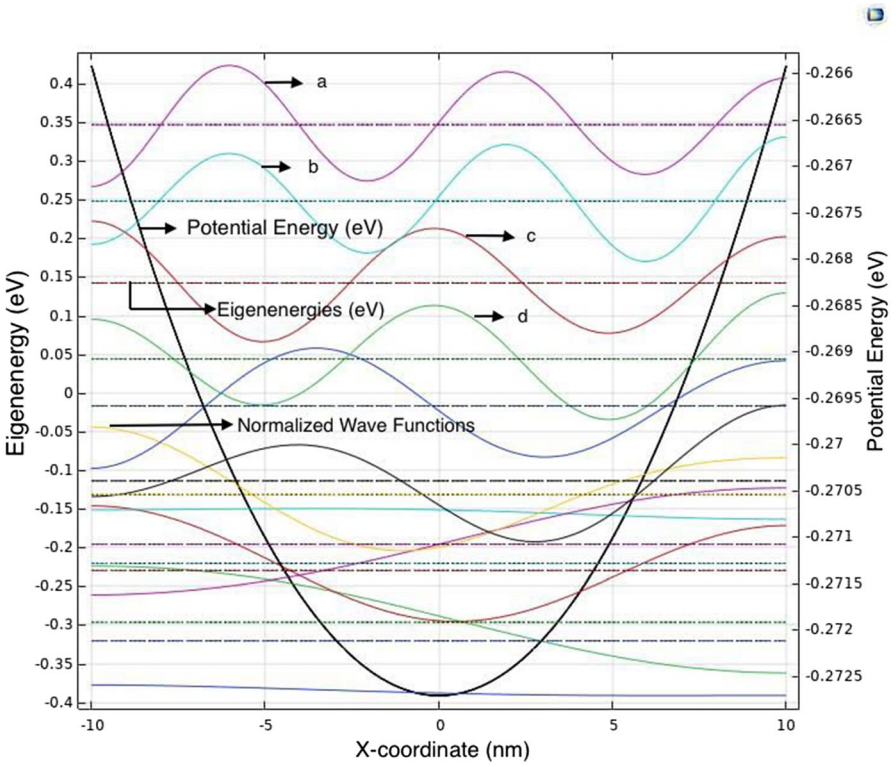


Fig. 2 Energy split sub-bands, wave function and applied potential on the quantum wire in the presence of external crossed electromagnetic fields, Zeeman effect and Rashba SOI

To calculate the thermodynamic properties of a system, it is necessary to derive the partition function to describe that system. Using the eigenenergies, the partition function can be derived by the relation,

$$Z = \sum_i e^{-\beta E_i} \tag{13}$$

where $\beta = \frac{1}{k_B T}$, $k_B = 1.38 \times 10^{-23} J/K$ is the Boltzmann constant that relates the average kinetic energy of particles in a gas to its temperature, and the summation is done for all the energy levels found using the Eqs. (10, 12).

We obtain the thermodynamic properties in terms of the partition function as follows [29]:

i. Entropy:

$$S = k_B log Z - k_B \beta \frac{\partial log Z}{\partial \beta} \tag{14}$$

Entropy refers to a thermodynamic quantity that measures the degree of disorder or randomness in a system at the nanoscale. It is a fundamental concept in statistical mechanics and provides insights into the arrangement and distribution of particles or molecules within the nanomaterial.

ii. Magnetic entropy:

$$\Delta S = S(B \neq 0, T) - S(B = 0, T) \tag{15}$$

Magnetic entropy represents the difference between entropy in the presence of some magnetic field and entropy in the absence of any magnetic field. It refers to the phenomenon where the temperature of a material changes in response to a magnetic field variation.

iii. Mean Energy:

$$U = -\frac{\partial logZ}{\partial \beta} \tag{16}$$

Mean energy is the slope of the log of the partition function Z vs β. It represents the average energy per particle or per degree of freedom in a system. Understanding the mean energy helps in characterizing the energy distribution, thermal behavior, and heat transfer mechanisms.

iv. Specific Heat:

$$C = \frac{\partial U}{\partial T} = k_B \beta^2 \frac{\partial^2}{\partial \beta} logZ \tag{17}$$

Specific heat is the amount of heat energy required to raise the temperature of a substance per unit mass or per mole. In nanomaterials, specific heat can exhibit size-dependent behavior due to quantum confinement effects and altered phonon and electron transport properties.

v. Free Energy:

$$F = -k_B T logZ \tag{18}$$

Free energy represents the energy available to do work in a system. In the context of nanomaterials, studying free energy is crucial for determining their stability, phase transitions, and equilibrium states.

3 Results

Here, the thermodynamic properties, viz. entropy, mean energy, free energy, and specific heat, have been studied as a function of temperature. The Quantum Wire is subjected to an electric field with field strength, $F = 2.0 \times 10^6 V/m$, and a magnetic field, $B = 8T$. The average value of Lande's g-factor has been fixed as, $g = -0.44$ [34], and Rashba effect strength, $\alpha = 2.0 \times 10^{-11} eVm$ is considered.

Figure 3a represents that rise in the temperature enhances the entropy of the quantum wire under the influence of hydrostatic pressure. At a high range of temperature,

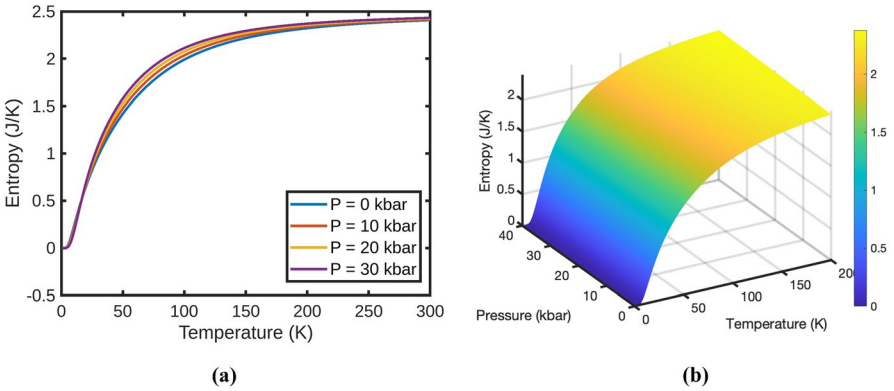


Fig. 3 **a** Entropy as a function of Temperature for different values of hydrostatic pressure and, **b** 3-dimensional plot of Entropy as a function of temperature and pressure

entropy reaches a constant value. For low temperatures, there is not much difference between the entropy values for different pressure values. Also, entropy is very close to zero in this temperature range (~5 K) as only the lowest energy levels are occupied and there is not enough energy available for transition. As temperature rises, sufficient energy is acquired for transition which is seen as an increase in the entropy. The particular values of temperature where the increase starts depend on various factors such as material of the wire, energy band gaps. Moreover, the increment in pressure causes a decrease in the dimensions of the quantum wire as a consequence, the volume of the system decreases. This results in lesser spacing among the energy levels thus making transitions more probable. Therefore around 50 K, the figure clearly shows that the entropy starts increasing with hydrostatic pressure. Furthermore, entropy is a degree of randomness or disorder in the system, increased transition probability results in an increase in entropy. A general trend of increase in entropy with an increase in pressure can be seen in Fig. 3b more clearly.

Figure 4a shows the variation of magnetic entropy with temperature; it shows a structure with a profound minimum at low temperature. After the minimum, the absolute value decreases and tends to zero with further increasing the temperature. From the figure, it becomes clear that on increasing the hydrostatic pressure, the absolute value of the magnetic entropy reduces.

From Eq. (6), the value of magnetic entropy is given as a difference of entropy at zero magnetic field and nonzero magnetic field. Figure 4b shows the entropy versus temperature at different pressures for both zero and nonzero magnetic fields. The solid lines represent magnetic field, $B = 8T$ and the dashed lines represent zero magnetic field, i.e., $B = 0$. A similar trend is observed for both the magnetic fields, but it is seen that for nonzero value of magnetic field, entropy is lesser. This can be owed to the fact that when the magnetic field is zero, no spin energy bands splitting takes place and thus, the energy levels are degenerate. This results in more transition probability and thus higher value of entropy. On the other hand, when a nonzero magnetic field is introduced, energy levels are split due to SOI effects, thus

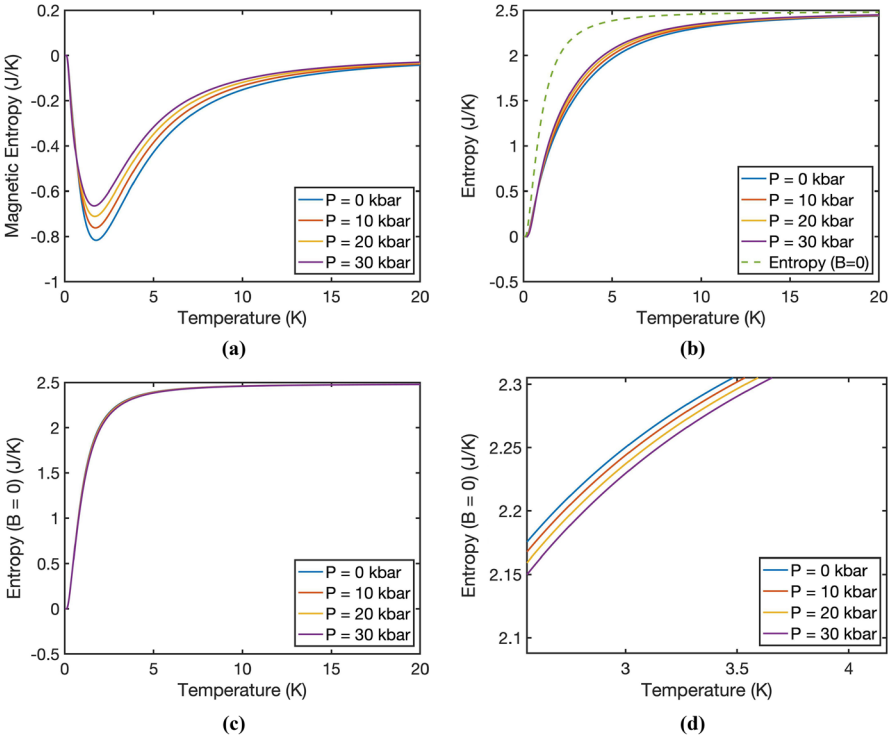


Fig. 4 **a** Magnetic entropy as a function of temperature at different values of hydrostatic pressure. **b** Entropy as a function of temperature for different values of hydrostatic pressure. For zero and nonzero magnetic fields. Pressure for green dashed line = 20 kbar. The solid lines represent $B = 8T$. **c** Entropy as a function of temperature with zero magnetic field. **d** Entropy as a function of temperature with zero magnetic field (zoomed in)

transitions become less probable due to difference in energy levels hence decreasing the entropy.

For finite B, entropy is found to increase with pressure. This can be understood by the compressing effect of pressure. The pressure reduces energy level spacing thus making transitions more probable thus increasing entropy. For $B = 0$, there is no Zeeman effect. Energy level splitting arises only due to spin coupling with momentum, i.e. Rashba coupling. In this case, the effect of pressure results in a decrease in the number of energy levels that are accessible thus a decrease in entropy is seen with an increase in pressure.

Moreover, on increasing the hydrostatic pressure, the value of entropy for nonzero magnetic fields increases. In Fig. 4c it is observed that for a zero magnetic field, there is not much difference between entropies at different pressures. A closer look in Fig. 4d reveals that for increase in hydrostatic pressure, the value of entropy decreases.

The above observations show that as pressure increases the difference between the entropy values at zero and nonzero magnetic fields decreases. This confirms

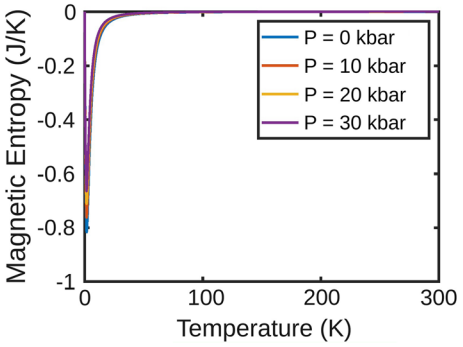


Fig. 5 Magnetic entropy v/s temperature extended to room temperature

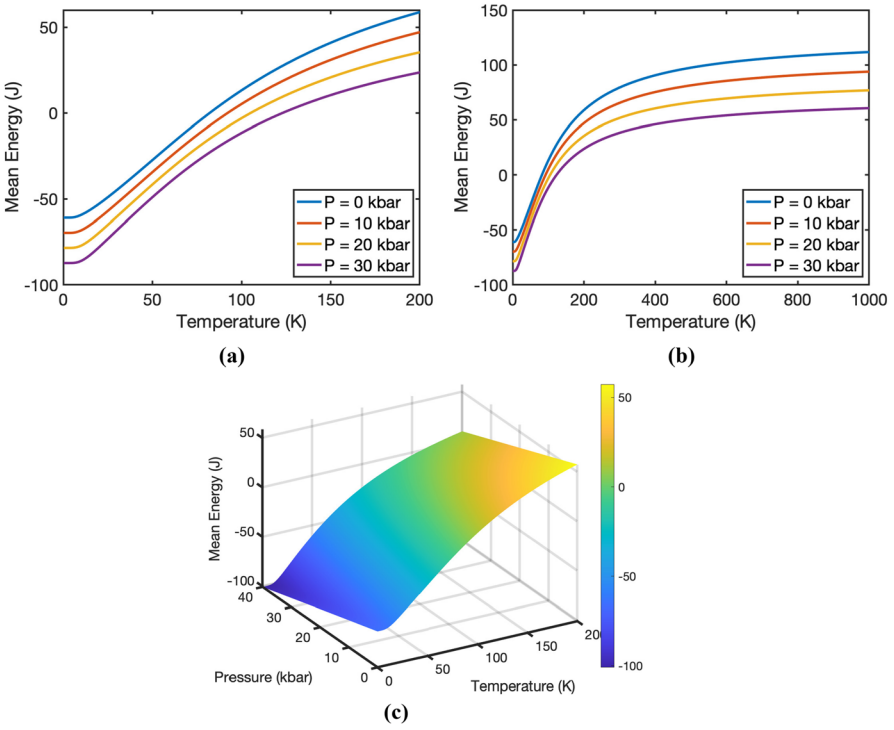


Fig. 6 Mean energy as a function of **a** low temperature, **b** high-temperature range for various values of hydrostatic pressure, and **c** 3-dimensional plot of mean energy as a function of temperature and pressure

that on increasing the value of hydrostatic pressure, the absolute value of magnetic entropy decreases.

Figure 5 shows the behavior of magnetic entropy when extended to room temperature.

In Figure 6a, b mean energy increases monotonically with temperature. The increase is clearly seen at a lower temperature in Fig. 6a. At a high range of

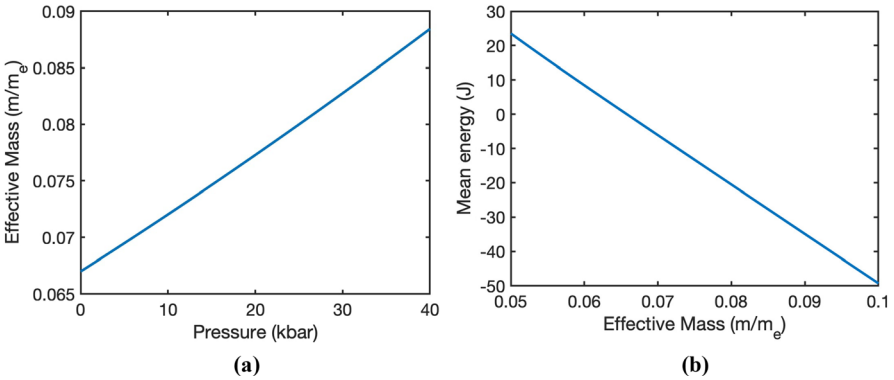


Fig. 7 **a** Effective mass dependence on hydrostatic pressure. **b** Mean energy v/s effective mass at $T = 80$ K

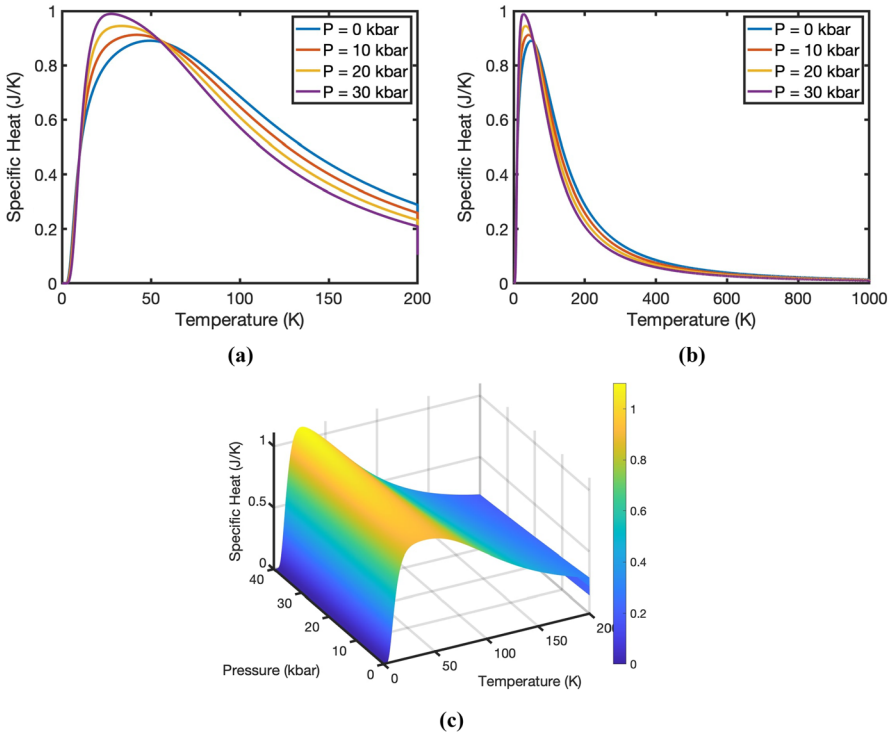


Fig. 8 Specific heat v/s temperature **a** low-temperature and, **b** high-temperature range at different values of hydrostatic pressure, **c** 3-dimensional plot of specific heat as a function of temperature and pressure

temperatures (~ 500 K), the increase slows down represented by Fig. 6b. Moreover, effective mass increases with the hydrostatic pressure and the mean energy reduces with increasing effective mass as shown in Fig. 7a and b, respectively. At very low

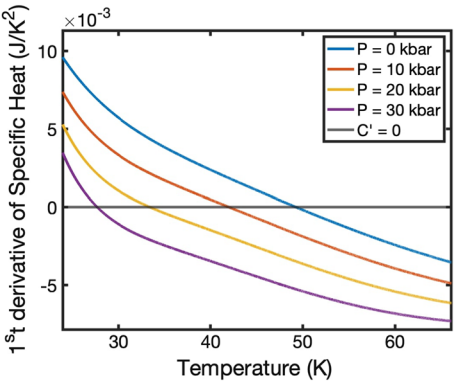


Fig. 9 First derivative of specific heat v/s temperature for different values of hydrostatic pressure

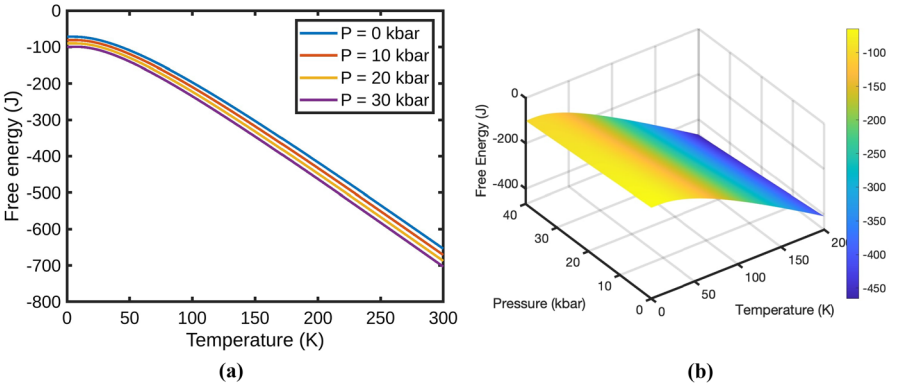


Fig. 10 **a** Free energy v/s temperature at different values of hydrostatic pressure and, **b** 3-dimensional plot of free energy as a function of temperature and pressure

temperatures (~5 K), mean energy is seen to be constant, this observation can be owed to the fact that at low temperatures only the lowest energy states are occupied and enough energy is not available for transition to higher states. As temperature increases, more energy is acquired and higher energy levels get populated. Thus, at a fixed temperature, mean energy is found to decrease with an increment in hydrostatic pressure.

Specific heat versus temperature, for both low- and high-temperature ranges, is shown in Fig. 8a, b. It can be easily deduced from Fig. 8a that specific heat increases with a rise in temperature at a low range of temperature and reaches a peak at 20–50 K. When the temperature is sufficiently high, all the highest energy levels are populated and further increasing the temperature does not increase the mean energy considerably (Fig. 6). Thus the gradient of mean energy starts to decrease; hence specific heat decreases. This can be seen as the peak in Fig. 9. After that, the specific heat gradually decreases with temperature. With the increment in pressure, the gap between the two energy levels reduces, which increases the transition probability.

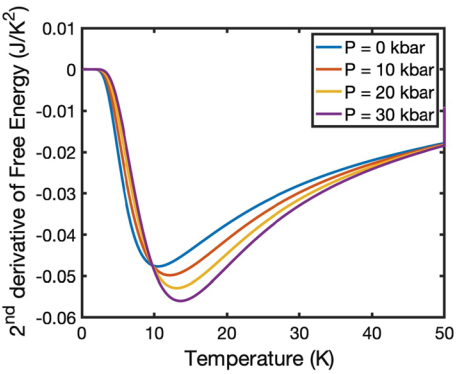


Fig. 11 Second derivative of free energy v/s temperature for different values of hydrostatic pressure

As a result, in Fig. 8b, the peak of the curve shifts towards low temperature with an increase in hydrostatic pressure.

Figure 9 shows how the first derivative of specific heat changes its sign. Here, the first derivative is used as a mathematical tool to investigate the curve of specific heat. The change in the sign of the first derivative indicates the peak of the corresponding curve. As pressure increases, the zero crossing of the first derivative happens at a lower temperature as a consequence, the peak of the curve shifts towards lower temperatures on increasing pressure.

Moreover, the free energy shows a similar trend to mean energy in terms of pressure, i.e. it decreases with an increase in pressure for a fixed value of temperature (Fig. 10). At temperatures lower than 20 K, free energy is nearly constant w.r.t. temperature, after which it starts to decrease rapidly with temperature. Figure 11 shows the second derivative of free energy (again used as a mathematical tool to investigate the curve of free energy) which shows that the second derivative reaches a minimum before starting to increase and then settles to a constant value. For higher values of hydrostatic pressure, the minimum is reached at a higher value of temperature as compared to lower values of pressure. This means that free energy remains nearly constant till a higher temperature value for higher pressure.

4 Conclusion

In this paper, the thermodynamic properties of a GaAs quantum wire have been investigated theoretically under the effect of hydrostatic pressure using a pressure-dependent effective mass approximation. The model considers a GaAs quantum wire under the influence of a crossed electromagnetic field with the inclusion of Rashba SOI which is controlled by an external gate voltage. The numerically obtained results show variations in thermodynamic properties against temperature at different pressures. At low temperatures, entropy, mean energy, and specific heat don't show much difference in values for different pressures. Entropy

is found to increase with increasing pressure, whereas mean energy and free energy are found to decrease with pressure. Magnetic entropy shows a minimum structure with respect to temperature and specific heat shows a structure with a maximum, where the peak position is found to be dependent on pressure. With increasing pressure, the peak shifts towards lower temperatures. Relating to the behavior of entropy, as pressure increases, the absolute value of magnetic entropy is decreased. These behaviors can be explained by energy level splitting due to Rashba SOI and the compressing effect of pressure which decreases the spacing between energy levels.

Author Contributions PK, SA, and YG contributed to Methodology, Software, Validation, Formal analysis, Data Curation, Writing—Original Draft, Conceptualization, Formal analysis, Writing—Review & Editing, and Visualization.Priyanka and RS contributed to Resources, Investigation, Supervision, Writing—Review & Editing, and Project administration.

Declarations

Competing interest The authors declare no competing interests.

References

1. C.M. Lieber, One-dimensional nanostructures: chemistry, physics & applications. Solid State Commun. **107**, 607–616 (1998). https://doi.org/10.1016/s0038-1098(98)00209-9
2. G.A. Prinz, Magnetoelectronics. Science **282**, 1660–1663 (1998). https://doi.org/10.1126/science.282.5394.1660
3. Y. Xia, P. Yang, Y. Sun, Y. Wu, B. Mayers, B. Gates, Y. Yin, F. Kim, H. Yan, One-dimensional nanostructures: synthesis, characterization, and applications. Adv. Mater. **15**, 353–389 (2003). https://doi.org/10.1002/adma.200390087
4. A. Martinez-Gil, A. Rota, T. Maroutian, B. Bartenlian, P. Beauvillain, E. Moyen, M. Hanbücken, Nano-patterned silicon surfaces for the self-organised growth of metallic nanostructures. Superlattices Microstruct. **36**, 235–243 (2004). https://doi.org/10.1016/j.spmi.2004.08.006
5. A.L. Fernandes Cauduro, L.H. Hess, D.F. Ogletree, J.W. Schwede, A.K. Schmid, Tailoring low energy electron absorption via surface nano-engineering of cesiated chromium films. Appl. Phys. Lett. **115**, 071602 (2019). https://doi.org/10.1063/1.5099115
6. S. Axelsson, Ee.B. Campbell, L.M. Jonsson, J. Kinaret, S.W. Lee, Y.W. Park, M. Sveningsson, Theoretical and experimental investigations of three-terminal carbon nanotube relays. New J. Phys. **7**, 245–245 (2005). https://doi.org/10.1088/1367-2630/7/1/245
7. C. Somaschini, S. Bietti, N. Koguchi, S. Sanguinetti, Coupled quantum dot–ring structures by droplet epitaxy. Nanotechnology **22**, 185602 (2011). https://doi.org/10.1088/0957-4484/22/18/185602
8. M. Law, D.J. Sirbuly, J.C. Johnson, J. Goldberger, R.J. Saykally, P. Yang, Nanoribbon waveguides for subwavelength photonics integration. Science **305**, 1269–1273 (2004). https://doi.org/10.1126/science.1100999
9. J.D. Sau, S. Tewari, S. Das Sarma, Universal quantum computation in a semiconductor quantum wire network. Phys. Rev. A (2010). https://doi.org/10.1103/physreva.82.052322
10. M. Sundaram, S.A. Chalmers, P.F. Hopkins, A.C. Gossard, New quantum structures. Science **254**, 1326–1335 (1991). https://doi.org/10.1126/science.254.5036.1326
11. J. Adamowski, A. Kwaśniowski, B. Szafran, LO-phonon-induced screening of electron–electron interaction in D-centres and quantum dots. J. Phys.: Condens. Matter **17**, 4489–4500 (2005). https://doi.org/10.1088/0953-8984/17/28/008

12. M. Ciurla, J. Adamowski, B. Szafran, S. Bednarek, Modelling of confinement potentials in quantum dots. Physica E **15**, 261–268 (2002). https://doi.org/10.1016/s1386-9477(02)00572-6
13. W. Xie, Investigation of D-centers confined by spherical quantum dots. Physica Status Solidi (b). **245**, 101–105 (2008). https://doi.org/10.1002/pssb.200743116
14. D.C. Banerjee, Review synthesis of conducting nanowires. J. Mater. Sci. **37**, 4261–4271 (2002). https://doi.org/10.1023/A:1020663731437
15. E.I. Rashba, Properties of semiconductors with an extremum loop. I. Cyclotron and combinational resonance in a magnetic field perpendicular to the plane of the loop. Soviet Phys. Solid State. **2**, 1109 (1960)
16. G. Dresselhaus, Spin-orbit coupling effects in zinc blende structures. Phys. Rev. **100**, 580–586 (1955). https://doi.org/10.1103/physrev.100.580
17. G. Engels, J. Lange, Th. Schäpers, H. Lüth, Experimental and theoretical approach to spin splitting in modulation-dopedInxGa1−xAs/InP quantum wells for B→0. Phys. Rev. B **55**, R1958–R1961 (1997). https://doi.org/10.1103/physrevb.55.r1958
18. T. Koga, J. Nitta, T. Akazaki, H. Takayanagi, Rashba spin-orbit coupling probed by the weak antilocalization analysis inInAlAs/InGaAs/InAlAs quantum wells as a function of quantum well asymmetry. Phys. Rev. Lett. (2002). https://doi.org/10.1103/physrevlett.89.046801
19. Priyanka, R. Sharma, M. Kumar, Effect of hydrostatic pressure and temperature on the ballistic conductance under the influence of Rashba spin-orbit coupling, Physica B: Condensed Matter. 648 (2023) 414402. https://doi.org/10.1016/j.physb.2022.414402.
20. A.J. Read, R.J. Needs, K.J. Nash, L.T. Canham, P.D.J. Calcott, A. Qteish, First-principles calculations of the electronic properties of silicon quantum wires. Phys. Rev. Lett. **69**, 1232–1235 (1992). https://doi.org/10.1103/physrevlett.69.1232
21. R. Khordad, Simultaneous effects of electron–electron interactions, Rashba spin-orbit interaction and magnetic field on susceptibility of quantum dots. J. Magn. Magn. Mater. **449**, 510–514 (2018). https://doi.org/10.1016/j.jmmm.2017.10.085
22. S.A. Tarasenko, N.S. Averkiev, Interference of spin splittings in magneto-oscillation phenomena in two-dimensional systems. J. Exp. Theor. Phys. Lett. **75**, 552–555 (2002). https://doi.org/10.1134/1.1500719
23. S. Dahiya, S. Lahon, R. Sharma, Effects of temperature and hydrostatic pressure on the optical rectification associated with the excitonic system in a semi-parabolic quantum dot. Physica E: Low-Dimens. Syst. Nanostruct. **118**, 113918 (2020). https://doi.org/10.1016/j.physe.2019.113918
24. R. Khordad, Optical properties of quantum wires: Rashba effect and external magnetic field. J. Lumin. **134**, 201–207 (2013). https://doi.org/10.1016/j.jlumin.2012.08.047
25. M. Kumar, S. Lahon, P.K. Jha, M. Mohan, Energy dispersion and electron g-factor of quantum wire in external electric and magnetic fields with Rashba spin orbit interaction. Superlattices Microstruct. **57**, 11–18 (2013). https://doi.org/10.1016/j.spmi.2013.01.007
26. R. Khordad, H.R. Rastegar Sedehi, Low temperature behavior of thermodynamic properties of 1D quantum wire under the Rashba spin-orbit interaction and magnetic field. Solid State Commun. **269**, 118–124 (2018). https://doi.org/10.1016/j.ssc.2017.10.018
27. Y. Karaaslan, B. Gisi, S. Sakiroglu, E. Kasapoglu, H. Sari, I. Sokmen, Rashba spin-orbit coupling effects on the optical properties of double quantum wire under magnetic field. Superlattices Microstruct. **93**, 32–39 (2016). https://doi.org/10.1016/j.spmi.2016.02.043
28. D. Najafi, B. Vaseghi, G. Rezaei, R. Khordad, Thermodynamics of mono-layer quantum wires with spin-orbit interaction. Eur. Phys. J. Plus (2018). https://doi.org/10.1140/epjp/i2018-12102-3
29. R. Khordad, H.R.R. Sedehi, Thermodynamic properties of a double ring-shaped quantum dot at low and high temperatures. J. Low Temp. Phys. **190**, 200–212 (2017). https://doi.org/10.1007/s10909-017-1831-x
30. R. Khordad, Ar. Firoozi, H.R.R. Sedehi, Simultaneous effects of temperature and pressure on the entropy and the specific heat of a three-dimensional quantum wire: tsallis formalism. J. Low Temp. Phys. **202**, 185–195 (2020). https://doi.org/10.1007/s10909-020-02536-w
31. P. Başer, I. Altuntas, S. Elagoz, The hydrostatic pressure and temperature effects on hydrogenic impurity binding energies in GaAs/InxGa1-xAs/GaAs square quantum well. Superlattices Microstruct. **92**, 210–216 (2016). https://doi.org/10.1016/j.spmi.2015.12.010
32. P. Baser, H.D. Karki, I. Demir, S. Elagoz, The hydrostatic pressure and temperature effects on the binding energy of magnetoexcitons in cylindrical quantum well wires. Superlattices Microstruct. **63**, 100–109 (2013). https://doi.org/10.1016/j.spmi.2013.08.021

33. S. Debald, B. Kramer, Rashba effect and magnetic field in semiconductor quantum wires. Phys. Rev. B (2005). https://doi.org/10.1103/physrevb.71.115322
34. A.M. Babayev, Ö. Mercan, S. Tez, Electron Lande g-factor in GaAs/Al Ga1−As quantum wires. Physica E **41**, 345–348 (2009). https://doi.org/10.1016/j.physe.2008.06.030

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 27 2023

Impact of wind effects on high-rise buildings having varying cross sections FREE

Devesh Kasana; Dhawal Tayal; Dhruv Choudhary; S. Anbukumar; Ritu Raj; Rahul Kumar Meena

Check for updates

AIP Conference Proceedings 2721, 040030 (2023)
https://doi.org/10.1063/5.0153990

View Online
Export Citation
CrossMark

10 August 2023 06:51:42

500 kHz or 8.5 GHz?
And all the ranges in between.
Lock-in Amplifiers for your periodic signal measurements

Find out more
Zurich Instruments

AIP Publishing

Impact of Wind Effects on High-rise Buildings Having Varying Cross Sections

Devesh Kasana[1], Dhawal Tayal[1] , Dhruv Choudhary[1], S. Anbukumar[1], Ritu Raj[1] and Rahul Kumar Meena[1a]

[1]Department of Civil Engineering, Delhi Technological University, New Delhi, 110042, India

a) Corresponding Author: rahul.08dtu@gmail.com

Abstract. Wind is regarded as one of the significant horizontal forces that strongly affects how high-rise constructions structures respond when they are designed. Population growth has increased demand for tall structures, which is growing daily. Numerous international standards provide pressure coefficients and force coefficients, which are used to calculate the wind load on such structures. These international standards, however, provide details on structures with regular shaped building, like squares and rectangles. With the development of technology, composite plan shape buildings like pentagon, hexagon, etc. have been given consideration by many architects while keeping in mind the building's aesthetics. Since there is no data on these buildings, it is necessary to do CFD or wind tunnel testing to study the effects of wind on these structures. ANSYS: CFD (Computation Fluid Dynamics) to predict wind reactions is rapidly gaining popularity. In the current research, two structure model with the plan shapes of a square and a pentagon have been taken into consideration. On each face of the square and pentagon buildings, the average value of the coefficient of pressure Cp is computed. Results are compared to international standards codes. The lever side is discovered to have negative pressure, and the pressure contours for the side faces are identical. A recirculation zone is shown by vertical flow patterns to be one-third of the building's height. While horizontal flow patterns show the development of a vortex in the building's wake.

Keywords. high-rise building, CFD, ANSYS, pressure contours, flow patterns

10 August 2023 06:51:42

INTRODUCTION

Building taller structures is the biggest development trend the globe is witnessing right now. A greater number of high-rise structures with unusual shapes have been constructed as a result of advancements in construction techniques, technology, and materials. High-rise structures must be effectively designed for wind loads because they are particularly vulnerable to them. Results of wind load for typical plan-shaped structures can be found in a number of international standards, including IS: 875 (Part III) - 2015. Wind tunnel testing and ANSYS: CFD technologies are used to study the influence of wind on high-rise buildings. Nowadays, ANSYS: CFD simulations are carried out to get wind forces in complex plan-shaped structures.

Numerous experimental and numerical studies have been done to date to investigate the effects of wind-driven forces on a variety of tall building models with unusual geometrical configurations. Many researchers have conducted wind tunnel experiments to assess the effects of wind on such structures. For instance, Bhattacharyya and Dalui [1] tested a "E"-shaped building in the wind tunnel; Nagar et al. [2] modelled the effects of wind on a "H" and a square building; Raj et al. [3] conducted experiments on a "Plus" shape building; Carassale et al. [4] researched the effect on a cylindrical square building; Yukio Tamura [5] examined triangular; Tanaka et al. studied improved square shaped building [6]; He and Song investigated the response of TTU building [7]; Pal et al. [8] evaluated drag and lift force coefficients on fish and square plan shape models ; Pal and Raj [9] studied the full blockage condition for various plan shapes of high-rise building models; Pal et al. [10] studied the effect of twin building models on a remodeled triangle shape model; and Nagar et al. [11] studied the effect on two "Plus" shape model. Following are a few significant findings from these studies: For isolated buildings, the drag and lift force coefficients depend on the location and geometric design of the building model, and the pressure contour is symmetric. Experiments in the wind tunnel take a lot of time and money. As a result, ANSYS: CFD approaches for analysing the impact of wind on tall buildings are growing in popularity.

The wind effects are provided by ANSYS: CFD approach is swift in time for different shapes of high-rise buildings. It uses meshing to break down the building geometry into smaller pieces, and various types of research have been conducted on a variety of shapes, such as the "H"-shaped building studied by Raj et al. [12]; the rectangular shape with

2nd International Conference on Futuristic and Sustainable Aspects in Engineering and Technology
AIP Conf. Proc. 2721, 040030-1–040030-11; https://doi.org/10.1063/5.0153990
Published by AIP Publishing. 978-0-7354-4576-5/$30.00

horizontal limbs studied by Paul and Dalui [13] and Okajima et al. [14]; the "Plus"-shaped building studied by Paul and Dalui [15]; the octagonal building studied by Hajra and Dalui [16]; Gaur and Raj studied on a square mode [17]; Meena et al. [18] studied rectangular shape buildings with different corner configurations; Goyal et al. [19] studied “Y” shape building; Raj et al. [20] looked at different cross-section building, Verma et al. [21] and Meena et al. [22] studied the cfd simulation on tall building. The simulation's main conclusions are that the mesh size has a substantial impact on the outcomes, that the height of the building model affects the pressure distribution, and that the exposure to wind affects the drag coefficients.

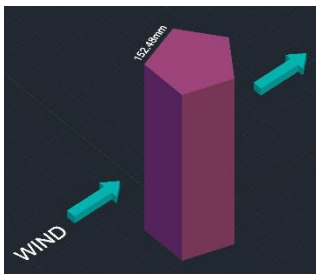
Using the computational fluid dynamics (CFD) method, the impact of wind on tall buildings with square and pentagonal cross sections is investigated. The outcomes of numerical modelling are evaluated in comparison to international standards. To obtain the streamlines or flow patterns around the building models and to obtain distinct flow characteristics, ANSYS: CFD approach is employed The pressure contour on various building faces is explored, and the building models horizontal and vertical streamlines are closely examined.

NUMERICAL MODEL DEVELOPMENT

In this study, pentagon and rectangular plan shaped buildings are analyzed by using the ANSYS CFX for CFD simulation at various wind incidence angle.

Geometric Modelling

In this study, Model A and Model P is designed in AutoCAD: 3D Modelling. Model P is a pentagon building, 600mm in height and 152.48mm in sides. Model A is a square building, 600mm in height and 200mm in sides. The isometric view of both buildings is shown in Fig. 1 and 2. The buildings are modelled using a 1:350 scale and the buildings are assumed as non-deformable. Both the buildings have an equal cross-sectional area of 40000mm^2.



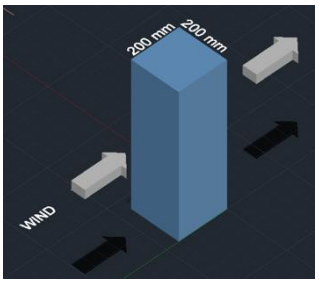


FIGURE 1. Model P **FIGURE 2.** Model A

Five steps make up the ANSYS analysis process: geometry, mesh, setup, solution, and result. To prevent flow recirculation, a rectangular domain is created in accordance with Standards and different research [18] with the borders placed 5H from the model's windward face (A), lateral faces (C & D), and roof and 15H from the leeward face (B). Let H be the height of the building model. Refer Fig. 3.

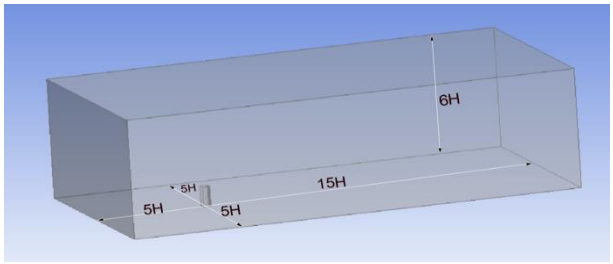


FIGURE 3. Domain used for simulation

10 August 2023 06:51:42

Meshing is the most crucial part of CFD simulation. A finer mesh may not give you perfect results, but an inadequate mesh will certainly affect the results. Meshing is of different shapes and sizes and provides valuable information on the flow characteristics in critical zones. Meshing divides the geometry into elements that can be used to discretize a domain. Meshing for the domain is done using the tetrahedral elements having an element size of 0.2 m, refer Fig. 4. Meshing Inflation is provided to simulate the flow easily around the building model's geometry. Meshing inflation with 15 layers is provided at the boundaries of Model A and Model P for allowing smooth flow and is represented in Fig. 5 and Fig. 6.

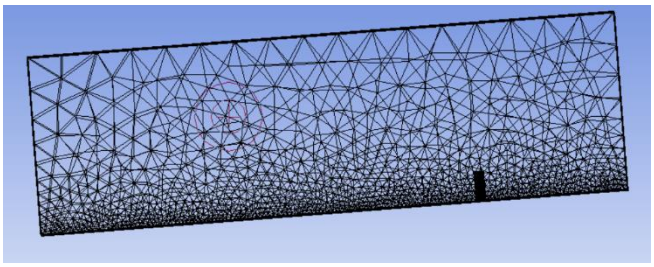


FIGURE 4. Domain meshing

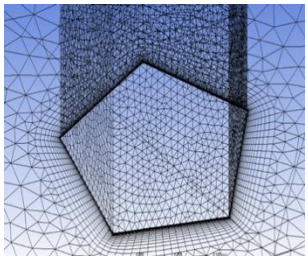


FIGURE 5. Model P Inflation layers

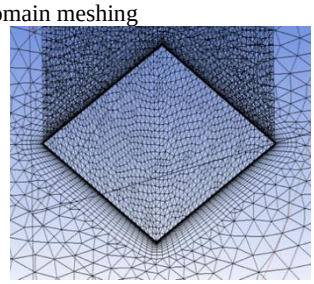


FIGURE 6. Model A Inflation layers

10 August 2023 06:51:42

Basic Equations

The study aims to find the fluid-structure interaction in terms of various streamlines and contours obtained as a result and validate the observations to conform with various international codes and standards. Many past studies were performed using the k- ε model. This model has a well-developed prediction capability and gives an accurate balance of reliability for general-purpose simulations. k-ε turbulence is a two-equation model that provides the solution by using two transport equations, that is, turbulent kinetic energy (k) and turbulent dissipation rate (ε).

The basic equations used to study the fluid flow problems are Navier-Stokes and continuity equations.

- Navier-Stokes equation is as follows:

$$\frac{\partial(\rho u_i)}{\partial t} = -\frac{\partial(\rho u_i u_j)}{\partial x_j} - \frac{\partial P}{\partial x_j} + \frac{\partial}{\partial x_j}[\mu(\frac{\partial_{u_i}}{\partial_{x_j}} + \frac{\partial_{u_j}}{\partial_{x_i}})] + F$$

- The standard k-ε model uses the following transport equations for the turbulence kinetic energy and turbulence dissipation:

$$\rho\frac{\partial k}{\partial t} + \rho\overline{u_i}\frac{\partial k}{\partial x_j} = \tau_{ij}\frac{\partial\overline{u_i}}{\partial x_j} - \rho\varepsilon + \frac{\partial}{\partial x_j}[(\mu + \frac{\mu_t}{\sigma_\varepsilon})\frac{\partial\varepsilon}{\partial x_j}]$$

- The momentum equation is as follows:

$$\frac{\partial(\rho U_i)}{\partial t} = -\frac{\partial(\rho U_i U_j)}{\partial x_j} - \frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j}[(\mu_{eff}(\frac{\partial_{U_i}}{\partial_{x_j}} + \frac{\partial_{U_j}}{\partial_{x_i}})] + S_M$$

CFD Pre-simulations

Power law is used for interpreting the results in ANSYS in set–up and solution steps. The inputs for the set–up step are as below. Refer Fig. 7.

- $\alpha = 0.147$
- $Z_0 = 1$ m
- $U_0 = 10$ m/s

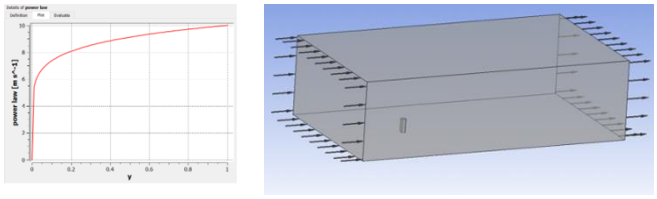


FIGURE 7. Set up in ANSYS CFX

Power law is used for the calculation of pressure.

$$P = U_{ref}\left(\frac{Z}{Z_{ref}}\right)^{\alpha}$$

The variable coefficient of pressure "Cp" is then found by,

$$C_p = \frac{P - P_o}{\frac{1}{2}\rho {U_{ref}}^2}$$

Where,

P is the pressure derived from the needed point, P_o is the reference height static pressure, ρ is the air density (1.225 kg/m^3), U_{ref} is the mean wind velocity at the building reference height(10 m/s).

The wind effect on Model A is analysed for 4 faces and the wind load effect on Model P is analysed for 5 faces. The faces are named as shown in Fig. 8.

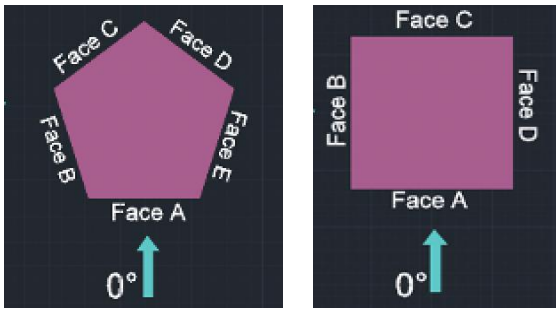


FIGURE 8. Faces representation for Models

10 August 2023 06:51:42

VALIDATION

For validating the study, Model A is chosen and analysed. Model A has a square cross-section having dimensions of 200mm x 200mm throughout the height of 600mm. After analysis using ANSYS: CFX, the coefficient of pressure, or Cp, values for each face A, B, C, and D classified similarly as discussed above were obtained. These values are listed in Table 1 below. The allowable values listed in various international standards are compared to it.

TABLE 1. C_p value for model a with acceptable C_p values under international standards

International standards	**Coefficient of pressure C_p**			
	Face A	***Face B***	***Face C***	***Face D***
IS: 875 (PART III) - 2015	+0.8	-0.8	-0.25	-0.8
ASCE/SEI 7-16	+0.8	-0.7	-0.5	-0.7
AS/NZS 1170.2.2011	+0.8	-0.65	-0.5	-0.65
ES/ISO 4354: 2012	+0.8	-0.65	-0.7	-0.65
CFD results for Model A	+0.68	-0.66	-0.29	-0.66

From Table 1, it may be observed that Cp values for faces A, B, C, and D varied by 15%, 17.5%, 16% and 17.5% respectively when compared with results by IS: 875 (Part III) – 2015.

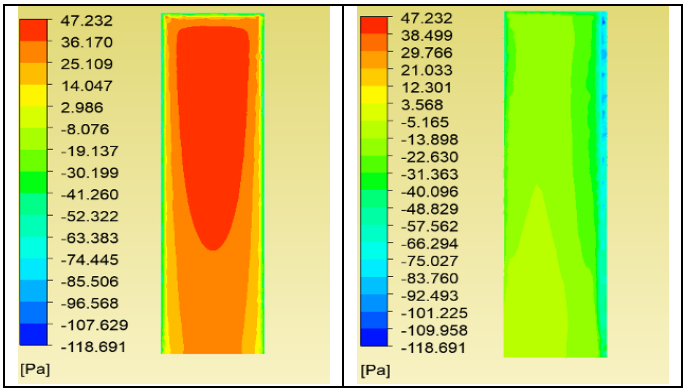
10 August 2023 06:51:42

RESULTS AND DISCUSSION

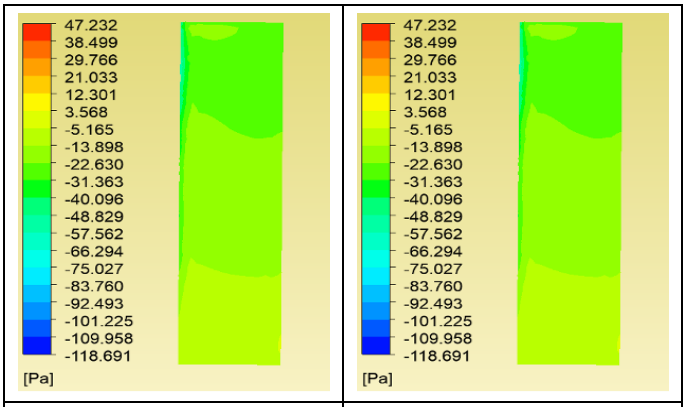
Pressure contours

Contours are used to indicate the change in constant pressure values over a surface. A contour plot with colour bands has discrete coloured regions while the display of a variable on a locator (such as a boundary) shows a finer range of colour detail. The following observations of pressure contours for all faces are listed below.

- Model P



Face A | Face B



Face C | Face D

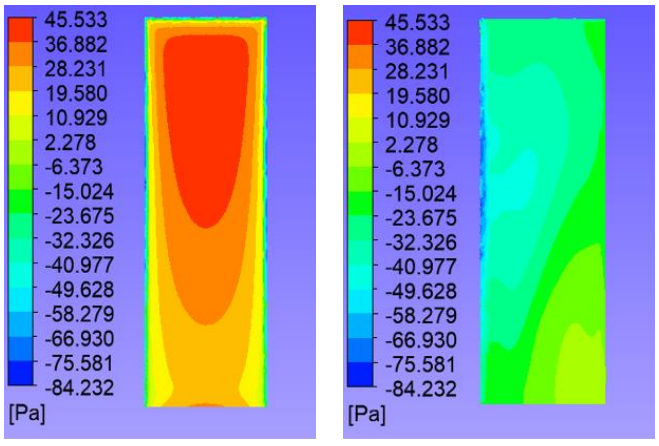
Face E | Top Face

FIGURE 9. Pressure contours for Model P

10 August 2023 06:51:42

The average, the coefficient of pressure C_P on faces A and B is 0.64, -0.35 on face C, -0.28 on face D, and -0.35 on face E. On face A, the highest positive pressure coefficient is observed 0.64. Similar negative pressure coefficient values of 0.35 are observed for Faces B and E. The negative pressure coefficient values for Faces C and D are also 0.28

- Model A



Face A | Face B

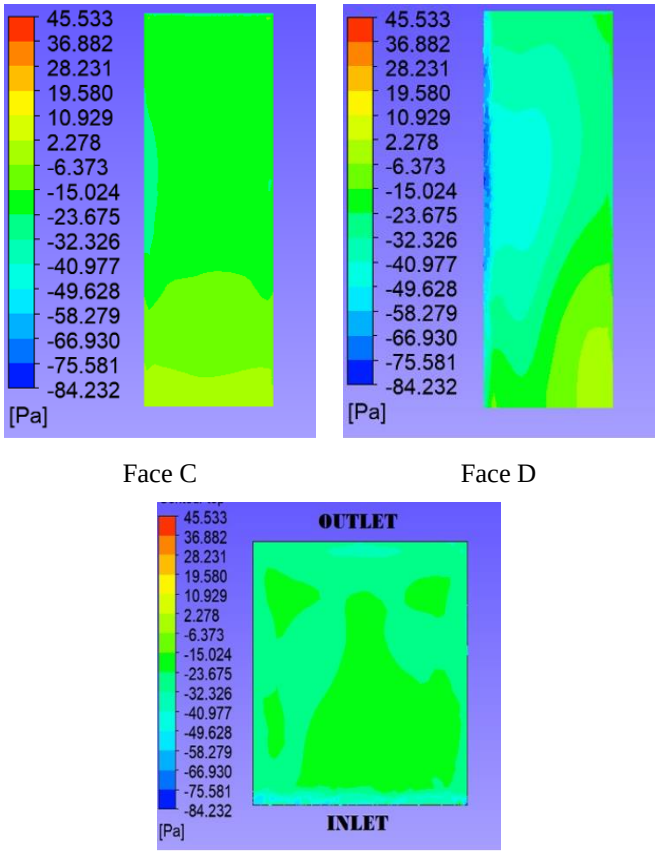


FIGURE 10. Pressure contours for Model A

The average value of the coefficient of pressure C_P on face A is 0.68, on face B is -0.66, on face C is -0.34, on face D is -0.66. The maximum positive pressure coefficient value is 0.68 occurred in face A. Faces B and D are getting similar negative pressure coefficient values of 0.66.

Streamlines or flow patterns

The wind in the present study is considered to be fluid. Streamline is an imaginary path drawn to flowing fluid. The vertical and horizontal streamlines are evaluated and represented in the following sub-sections.

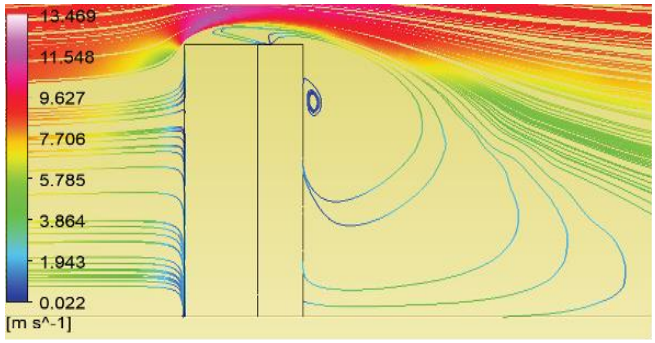


FIGURE 11. Vertical streamlines for model P

10 August 2023 06:51:42

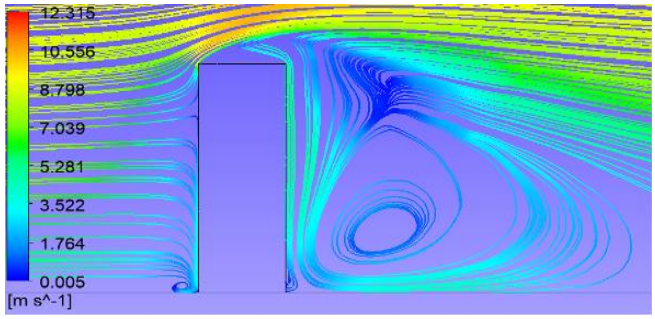
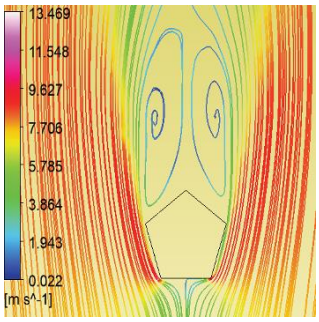
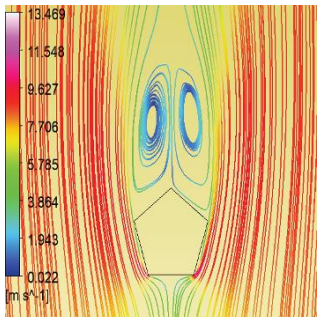


FIGURE 12. Vertical streamlines for model A

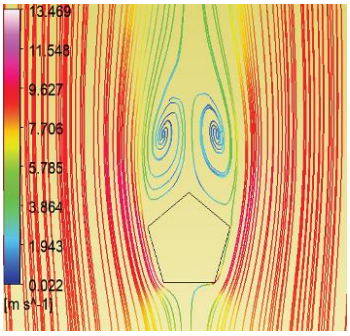
In both the models, it is observed that vertical streamlines from the bottom to 2/3 height of the building are changing their direction downward, and vertical streamlines from the top to 1/3 height of the building are changing their direction upward. It is observed that wind direction reverses in the wake region, creating a separation zone.



At 1/3 Plane



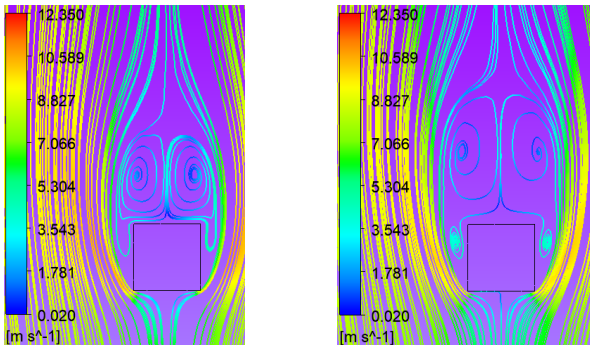
At mid Plane



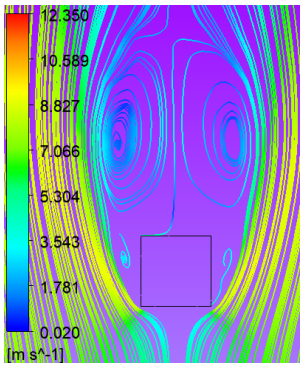
At 2/3 Plane

FIGURE 13. Horizontal streamlines for model P

10 August 2023 06:51:42



At 1/3 Plane At mid Plane



At 2/3 Plane

FIGURE 14. Horizontal streamlines for model A

In both models, recirculation zones are observed. Symmetrical horizontal streamlines are observed in both models. Two distinct vortices are visible in separation zone. At mid plane of square model two more small sized vortices are observed at faces B and D. When compared to the square model, the pentagon model's vortices are further distant from the back face. The pentagon's vortices are larger with respect top square model.

CONCLUSION

The wind is a great complex phenomenon because of the different flow situations developing from the interaction of wind with high-rise structures. The wind is composed of a multitude of eddies of different sizes and rotational characteristics carried along in a general stream of air moving relative to the earth's surface.

The complex nature of wind is due to its turbulent and gutsy nature that gives rise to many flow conditions upon interaction with tall buildings. Therefore, the study of wind flow characteristics on tall buildings is necessary for investigating the behaviour of these structures. The following conclusions may be drawn from this paper:

- International standards are used for analysing wind effects on tall buildings having simple shapes as cross–section only (square, Model A). Therefore, in this study complex shapes (pentagon, Model P) have been analysed by adopting CFD analysis and k - ε turbulence model in ANSYS: CFD model for interpreting the wind flow characteristics.
- As discussed in the validation section, the Cp values obtained for Model A were in agreement with those given by IS: 875 (Part III) – 2015 and other international codes. Therefore, Model P is also validated by consequence.
- Wind flow characteristics were determined for all faces of Model A and Model P in terms of pressure contours and vertical and horizontal streamlines.
- The maximum positive pressure coefficient value was maximum on face A of both models. In model A, face B and face D showed identical results. In model P, face B and face E showed identical results and face C and face D showed identical results.

10 August 2023 06:51:42

- Moreover, the Use of ANSYS: CFD for wind analysis in tall buildings is better than other popular methods like wind tunnel testing as it is less time-consuming and cost-effective.

ACKNOWLEDGEMENT

The authors are thankful to **Delhi Technological University** for providing the research facilities and funding to support this study.

REFERENCES

[1] B. Bhattacharyya and S. K. Dalui, "Investigation of mean wind pressures on 'E' plan shaped tall building," *Wind Struct. An Int. J.*, vol. 26, no. 2, pp. 99–114, 2018, doi: 10.12989/was.2018.26.2.099.

[2] S. K. Nagar, R. Raj, and N. Dev, "Experimental study of wind-induced pressures on tall buildings of different shapes," *Wind Struct. An Int. J.*, vol. 31, no. 5, pp. 441–453, 2020, doi: 10.12989/was.2020.31.5.431.

[3] R. Raj, S. Jha, S. Singh, and S. Choudhary, "Response analysis of plus shaped tall building with different bracing systems under wind load," *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 3, pp. 371–380, 2020, doi: 10.34218/IJARET.11.3.2020.032.

[4] L. Carassale, A. Freda, and M. Marrè-Brunenghi, "Experimental investigation on the aerodynamic behavior of square cylinders with rounded corners," *J. Fluids Struct.*, vol. 44, pp. 195–204, 2014, doi: 10.1016/j.jfluidstructs.2013.10.010.

[5] E. K. Bandi, Y. Tamura, A. Yoshida, Y. Chul Kim, and Q. Yang, "Experimental investigation on aerodynamic characteristics of various triangular-section high-rise buildings," *J. Wind Eng. Ind. Aerodyn.*, vol. 122, pp. 60–68, 2013, doi: 10.1016/j.jweia.2013.07.002.

[6] H. Tanaka, Y. Tamura, K. Ohtake, M. Nakai, and Y. Chul Kim, "Experimental investigation of aerodynamic forces and wind pressures acting on tall buildings with various unconventional configurations," *J. Wind Eng. Ind. Aerodyn.*, vol. 107–108, pp. 179–191, 2012, doi: 10.1016/j.jweia.2012.04.014.

[7] J. He and C. C. S. Song, "A numerical study of wind flow around the TTU building and the roof corner vortex," *J. Wind Eng. Ind. Aerodyn.*, vol. 67–68, pp. 547–558, 1997, doi: 10.1016/S0167-6105(97)00099-8.

[8] S. Pal, R. Raj, and S. Anbukumar, "Comparative study of wind induced mutual interference effects on square and fish-plan shape tall buildings," *Sadhana - Acad. Proc. Eng. Sci.*, vol. 46, no. 2, p. 27, 2021, doi: 10.1007/s12046-021-01592-6.

[9] S. Pal and R. Raj, "Evaluation of Wind Induced Interference Effects on Shape Remodeled Tall Buildings," *Arab. J. Sci. Eng.*, vol. 46, no. 11, pp. 11425–11445, 2021, doi: 10.1007/s13369-021-05923-x.

[10] S. Pal, R. Raj, and S. Anbukumar, "Bilateral interference of wind loads induced on duplicate building models of various shapes," *Lat. Am. J. Solids Struct.*, vol. 18, no. 5, 2021, doi: 10.1590/1679-78256595.

[11] S. K. Nagar, R. Raj, and N. Dev, "Proximity effects between two plus-plan shaped high-rise buildings on mean and RMS pressure coefficients," *Sci. Iran.*, vol. 0, no. 0, p. 28, 2021, doi: 10.24200/sci.2021.55928.4484.

[12] R. Raj, T. Rana, T. Anchalia, and U. Khola, "Numerical study of wind excited action on H Plan-shaped tall building," *Int. J. Emerg. Technol.*, vol. 11, no. 3, pp. 591–605, 2020.

[13] R. Paul and S. K. Dalui, "Shape optimization to reduce wind pressure on the surfaces of a rectangular building with horizontal limbs," *Period. Polytech. Civ. Eng.*, vol. 65, no. 1, pp. 134–149, 2021, doi: 10.3311/PPci.16888.

[14] A. Okajima, D. Yi, A. Sakuda, and T. Nakano, "Numerical study of blockage effects on aerodynamic characteristics of an oscillating rectangular cylinder," *J. Wind Eng. Ind. Aerodyn.*, vol. 67–68, pp. 91–102, 1997, doi: 10.1016/S0167-6105(97)00065-2.

[15] R. Paul and S. K. Dalui, "Prognosis of wind-tempted mean pressure coefficients of cross-shaped tall buildings using artificial neural network," *Period. Polytech. Civ. Eng.*, vol. 64, no. 4, pp. 1124–1143, 2020, doi: 10.3311/PPci.16311.

10 August 2023 06:51:42

[16] S. Hajra and S. K. Dalui, “Numerical investigation of interference effect on octagonal plan shaped tall buildings,” *Jordan J. Civ. Eng.*, vol. 10, no. 4, pp. 462–479, 2016.

[17] N. Gaur and R. Raj, “Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel,” *Ain Shams Eng. J.*, vol. 13, no. 1, 2022, doi: 10.1016/j.asej.2021.06.007.

[18] R. K. Meena, R. Raj, and S. Anbukumar, “Wind Excited Action around Tall Building Having Different Corner Configurations,” *Adv. Civ. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/1529416.

[19] P. K. Goyal, S. Kumari, S. Singh, R. K. Saroj, R. K. Meena, and R. Raj, “Numerical Study of Wind Loads on Y Plan-Shaped Tall Building Using CFD,” *Civ. Eng. J.*, vol. 8, no. 2, pp. 263–277, 2022, doi: 10.28991/CEJ-2022-08-02-06.

[20] R. RAJ, A. RUKHAIYAR, B. JAYANT, and K. DAHIYA, “CFD Analysis for wind ow characteristics of varying cross section tall building using ANSYS,” 2022.

[21] A. Verma, R. K. Meena, H. Dubey, R. Raj, and S. Anbukumar, “Wind Effects on Rectangular and Triaxial Symmetrical Tall Building Having Equal Area and Height,” vol. 2022, 2022

[22] R. K. Meena, R. Raj and S. Anbukumar. Effect of wind load on irregular shape tall buildings having different corner configuration. Sādhanā 47, 126 (2022). https://doi.org/10.1007/s12046-022-01895-2

10 August 2023 06:51:42

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Implementation of a Deepfake Detection System using Convolutional Neural Networks and Adversarial Training

Sahithi Bommareddy
Department of Software Engineering,
Delhi Technological University,
New Delhi, India
sahithibommareddy@gmail.com

Tanya Samyal
Department of Software Engineering,
Delhi Technological University,
New Delhi, India
tanyasamyal123@gmail.com

Sonika Dahiya
Department of Software Engineering,
Delhi Technological University,
New Delhi, India
sonika.dahiya@dtu.ac.in

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205614

***Abstract*— Due to the threat of spreading erroneous information, the employment of deep convolutional neural network (CNN) techniques has resulted in an increase in the number of altered images. A trustworthy and effective mechanism for identifying false photographs is required to solve this problem. In this study, we used the V4D architecture to analyze manipulated videos and concentrated on identifying DeepFake videos in three different contexts: I identifying all types of manipulations, (ii) identifying single manipulations, and (iii) identifying cross manipulation detection, which was used to assess the videos' authenticity. To do this, we used a variety of manipulation techniques and created algorithms to classify undiscovered manipulation methods.**

Keywords*— Deepfakes, Deepfake Detection, Media forensics, Facial Manipulation Detection, Convolutional Neural Networks, Manipulation techniques, Generative Adversarial Networks

I. INTRODUCTION

Deepfake technology is an emerging field of artificial intelligence that can create photoshopped films and images that look authentic. Advanced algorithms and machine learning techniques are utilized to make these films and photos, which result in realistic-looking content that can be used for everything from entertainment to propaganda. Although this technology can completely transform the media landscape, it also raises significant ethical and security problems [1].

Deepfake technology raises several issues, one of which is the possibility of malicious applications. Deepfake movies, for instance, can be used to slander people or groups, disseminate misleading information, and even affect political campaigns. Deepfake technology can also be used to fabricate sexual photographs or films that can be used for extortion or harassment [2]. The security and privacy of individuals, the integrity of public discourse, and democratic institutions are all gravely threatened by this.

The potential for deepfake technology to be exploited to undermine confidence in the media and other information sources is another worry [3]. Deepfake films may become more difficult to discern from real content as they get more lifelike, creating a general atmosphere of suspicion and uncertainty. When the correct information is essential, such as during natural catastrophes or political crises, this can be especially risky.

Deepfake technology could have a lot of advantages despite these worries. For instance, it can be applied to the film and entertainment industries to provide more authentic visual effects. It could also be employed for educational reasons, such as developing simulations and visual aids for instruction in science or medicine. Deepfake technology has the potential to improve many facets of our life with the right legislation and moral standards in place [4].

II. RELATED WORK

Researchers have created a range of methods to identify and stop deepfake content in response to these dangers. Some of these methods focus on examining the discrepancies and abnormalities found in deepfake media, like abnormal eye movements or distorted face characteristics. Some methods look for patterns in the data that separate authentic media from fake media using machine learning algorithms [5].

Another area of research has focused on developing new algorithms and tools for creating more realistic and convincing deepfake content. Some researchers have explored the use of generative adversarial networks (GANs) [6], a type of machine learning algorithm that involves two neural networks competing against each other to create realistic synthetic media. Other researchers have explored the use of neural style transfer, a technique that involves transferring the style and appearance of one image onto another image or video.

Deepfake films can be identified using a variety of methods. Among the most popular methods are:

- *Face Discrepancies:* Deepfake movies may contain small face discrepancies that are difficult to spot with the human eye but that computer algorithms can recognize. These discrepancies could be brought on by variations in skin tone, lighting, or facial features.
- *Abnormal Movements*: Deepfake films frequently have motion artifacts that are absent from authentic videos. These artifacts may include unusual motions or abnormal blurring around the margins of the subject.
- *Audio inconsistencies:* Deepfake movies may include audio that is out of rhythm with the motions of the subject. By examining the audio waveform and contrasting it with the video, this can be found.[7]
- *Eye blinking:* Artificial eye blinking patterns may be present in deepfake videos. By examining the frequency and timing of eye blinks, this can be discovered.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:08 UTC from IEEE Xplore. Restrictions apply.

Analysis of the video's metadata may disclose information about the source of the video, including the equipment or software that was used to capture or make it.

By examining patterns in the video and contrasting them with real footage, machine learning systems can be trained to recognize deep-fake videos.

The following are a few of the most popular algorithms:

A. *Deep Learning:* These techniques are frequently employed in the detection of deep fakes. To find patterns and discrepancies that can be utilized to discriminate between real and false videos, they entail training deep neural networks on enormous datasets of real and fraudulent videos.[8]

B. *CNNs (Convolutional Neural Networks*): They are a subset of deep learning algorithms that can be applied to the analysis of images and videos. They frequently analyze the visual elements of the video and look for any discrepancies to uncover deep fakes. [9]

C. *RNNs (Recurrent Neural Networks):* They are a different class of deep learning algorithms that can be applied to the identification of deepfakes. They can be used to identify discrepancies in the audio of deepfake films since they are particularly effective at understanding speech and audio patterns.

D. *GANs (Generative Adversarial Networks):* Deep learning algorithms called GANs are used to produce deepfake videos. But, by examining the patterns and discrepancies in the produced videos, they can also be used for deepfake detection.

E. *Face Recognition:* By examining the facial features of the subject in the video and contrasting them with known real and fake videos, face recognition algorithms can be used to detect deep fakes.

F. *Image Forensics:* By examining the metadata, compression artifacts, and other digital traces left behind in the movie, image forensic techniques can be used to identify deepfake videos.

There are various previously employed types of algorithms in the CNN algorithms:

A. *Xception:* For picture categorization and object recognition, Xception is a deep CNN architecture. By examining the visual characteristics of the video frames and looking for any irregularities or inconsistencies, it has been utilized for deepfake detection.[10]

B. *EfficientNet:* Another deep CNN architecture that has been applied to deepfake detection is EfficientNet. It is well-suited for deepfake detection because of its reputation for achieving excellent accuracy on picture classification tasks with only a small number of parameters.[11]

C. *ResNet:* It is a deep CNN architecture that has gained popularity for its capability to be able to train extremely deep neural networks by not having the problem of the vanishing gradient issue.[12] By examining the visual characteristics of the video frames and looking for any irregularities or inconsistencies, it has been utilized for deepfake detection.

D. *Inception:* Conceived initially for image classification applications, Inception is a deep CNN architecture. By examining the visual characteristics of the video frames and looking for any irregularities or inconsistencies, it has been utilized for deepfake detection.

E. *VGG:* For image classification problems, a deep CNN architecture called VGG is frequently employed. By examining the visual characteristics of the video frames and looking for any irregularities or inconsistencies, it has been utilized for deepfake detection.[13]

III. DATASET

A large-scale video dataset called FaceForensics++ was developed for deepfake detection studies. It is made up of more than 1,000 videos, amounting to more than 500GB of data, and is meant to mimic situations where deepfake videos might appear in the real world [14].

The videos in the dataset were produced using a range of deepfake methods, such as face swapping, facial reenactment, and puppeteering [15]. They are classified as "genuine" or "fake," with the fake videos further subdivided into four groups according to the process utilized to produce them.

In addition to the videos themselves, the FaceForensics++ dataset also contains ground truth labels that specify which movies are authentic and which are fraudulent, as well as metadata about the videos' resolution, frame rate, and compression level [16].

The FaceForensics++ dataset has been utilized in several studies and contests and has grown to be a well-liked resource for deepfake detection researchers. It has won accolades for being large, diverse, and realistic, which makes it ideal for developing and testing deepfake detection systems.

Researchers from the universities across Europe developed the FaceForensics++ dataset. The dataset was provided in 2019 to give academics working on deepfake detection access to a sizable, varied, and realistic dataset.

On the foundation of the 2018-released FaceForensics dataset, the newer version of the dataset was developed. A lesser number of films produced utilizing a range of deepfake methods, such as facial reenactment, face swapping, and neural rendering, made up the initial FaceForensics dataset.

The initial dataset was enhanced in several ways to produce the FaceForensics++ dataset. The collection was first expanded to include more than 1,000 videos, adding up to more than 500GB of data. Second, the authenticity of the videos was verified by ground truth labels that said whether they were real or phony, and if fraudulent, how they were made. Third, the dataset contains videos that have been compressed, blurred, and degraded to mimic the conditions that might be found in real-world circumstances.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:08 UTC from IEEE Xplore. Restrictions apply.

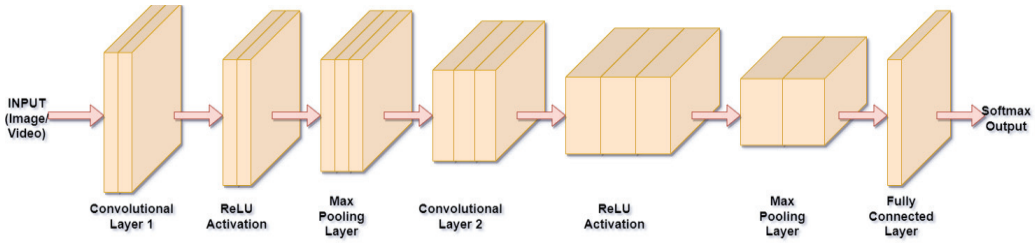


Fig. 1. Working diagram of C2D

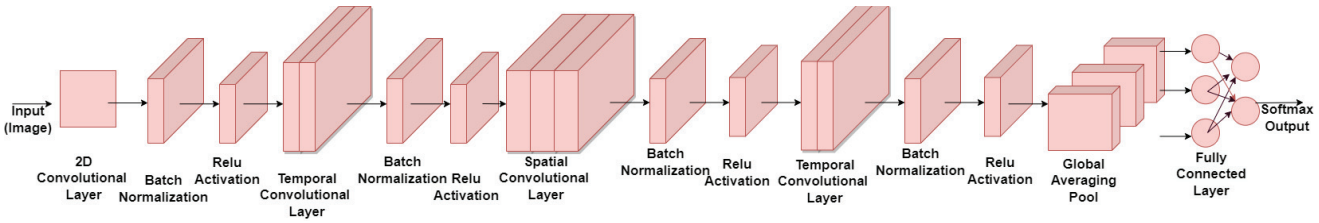


Fig. 2. Working diagram of R(2+1)D

IV. MODELS USED

In this section, we majorly talk about the models which we have used in order to test out our techniques.

A. C2D

For video classification applications, the C2D model is a convolutional neural network (CNN) architecture that is frequently employed. Convolutional Two Dimensional, or C2D, is the abbreviation used to refer to the network's use of 2D frames of video data [18].

A few layers of convolutional and pooling processes are commonly used in the construction of a C2D model, which is then followed by the fully connected layers for classification. The convolutional layers take the input frames and run them through a collection of learnable filters to extract features that are important for the classification task. The spatial resolution of the input data is decreased and the robustness of the features to small spatial fluctuations is increased by the pooling layers' down sampling of the feature maps created by the convolutional layers.

The C2D model often follows the convolutional layers with a few fully connected layers that carry out the actual classification [19], [20]. A vector of scores—one for each potential class—that may be read as the likelihood that the input video belongs to each class—the output of the final fully connected layer—is produced.

B. R(2+1)D

The 3D CNN architecture, which is frequently employed for video classification problems, has a version called the R(2+1)D model. The acronym "R(2+1)D" stands for "Residual (2+1)D," which alludes to the network's utilization of 2D and 1D convolutions in addition to its incorporation of residual connections.[21]

According to the R(2+1)D model, video data can be represented as a spatiotemporal volume having spatial dimensions (height and width), a temporal dimension (time), and a number of channels (e.g., RGB color channels). To capture both spatial and temporal aspects of the video data, the model combines 1D and 2D convolutions.

The 1D convolutions are performed along the temporal dimension, while the 2D convolutions are applied in the spatial dimensions of the input data. The model can capture temporal dynamics and learn features that are important for video classification tasks thanks to the usage of 1D convolutions.

The R(2+1)D model employs residual connections, a kind of skip connection that enables data to be transmitted straight from one layer to another, in addition to the usage of 2D and 1D convolutions. It has been demonstrated that residual connections make it simpler to train deep architectures, which enhances the performance of deep neural networks.[22]

C. CS-CNN

For image recognition tasks, the Channel-Separated Convolutional Neural Network (CS-CNN) deep learning model is used. It is intended to divide an image's color channels into distinct routes and process each pathway independently.[23]

Traditionally used convolutional neural networks process color channels in a single circuit. The CS-CNN model, on the other hand, divides each color channel (red, green, and blue) into its own pathway, enabling more effective and precise processing of color data.

The CS-CNN model is made up of several convolutional layers, each of which uses a collection of learnable filters to extract key characteristics from the input picture or feature map. The CS-CNN model produces different feature maps for each color channel by convolving each color channel with a unique set of filters.

Each color channel's output feature maps are then joined together and supplied into the following layer for additional processing. With this method, the model can learn representations of the input image that are more intricate by taking into consideration the distinctive properties of each color channel.

D. V4D

A variation of the well-liked ResNet architecture for deep learning models is V4D, commonly referred to as ResNet2. ResNet, an acronym for "Residual Network," was developed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:08 UTC from IEEE Xplore. Restrictions apply.

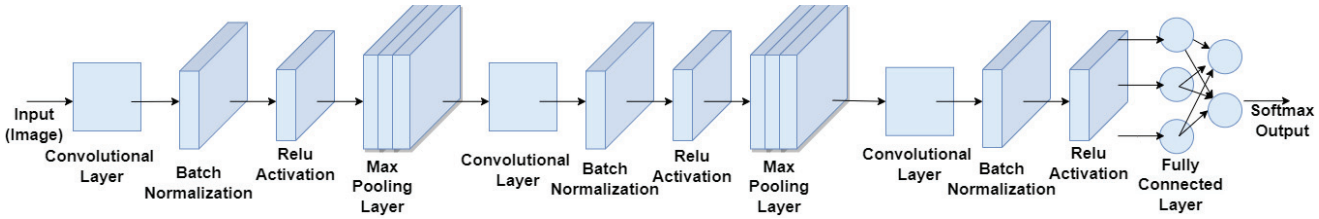


Fig. 3. Working diagram of CS-CNN

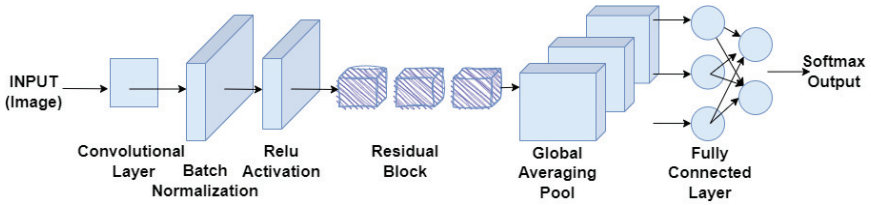


Fig. 4. Working diagram of V4D

by Microsoft researchers in 2015 and has since grown in popularity and influence in the field of computer vision.[24]

V4D (short for "Vision for Decisions") extends the ResNet architecture by adding a number of fresh elements and enhancements. Among the most significant modifications are:

Multi-path architecture: To enhance information flow and broaden the range of features accessible to downstream levels, V4D incorporates a number of parallel paths, each with a distinct level of abstraction.

Stochastic depth: During training, V4D randomly removes entire layers, preventing overfitting and enhancing generalization.

Shake-shake regularization is a novel type of regularization in V4D that increases generalization by introducing noise during the training phase.

In instances when the statistics of the entire batch might not be representative or in smaller batch sizes, V4D uses group normalization rather than batch normalization, which can be more productive in these circumstances.

V. EXPERIMENTAL SETUP

In our study, we implemented several ResNet-based models using PyTorch. These models, which used the action-based Kinetics 400 dataset to train them, included the C2D (ResNet50), R(2+1)D (ResNet50), V4D (ResNet2), and CSN (ResNet101) architectures. The FaceForensics++ dataset was used to adjust the neural network models once they had been initialized with weights. Using two NVIDIA TU102 [TITAN RTX] GPUs, we trained all of the models end-to-end using the Adam optimizer at a constant learning rate of 1e3 throughout all tests. The models needed pipelines with inputs that had 224 × 224 spatial dimensions and 64 video frames per frame during training. Based on the model architecture, we train the models which includes temporal subsampling, normalization, scaling, and image cropping. Each grade was identified using the Dual Shot Face Detector, and we then tried to make the region mentioned above in a larger scale by increase to ensure uniform coverage of the image and give the model more data.

VI. RESULTS

In this there was an experiment of the above listed architectures which are C2D, R(2+1)D, V4D and CSN on all 4 manipulation techniques. Their main goal is to learn and identify manipulation techniques. Apart from this there are individual techniques for each of the models and techniques for cross manipulation. Following that, we divided the data into one with various training, the testing, and validation groups using the technique described above.

A. Miscellaneous manipulation techniques

TABLE I. TRUE CLASSIFICATION RATES (TCR) ON FOUR MANIPULATION METHODS

Algorithm	Train and Test	TCR
Steg. Features + SVM [25]	FS, DF, F2F, NT	55.98
Cozzolino et al. [26]	FS, DF, F2F, NT	58.69
Bayar and Stamm [27]	FS, DF, F2F, NT	66.84
Rahmouniet al. [28]	FS, DF, F2F, NT	61.18
MesoNet [29]	FS, DF, F2F, NT	70.47
XceptionNet [30]	FS, DF, F2F, NT	81.0
3D ResNet [31]	FS, DF, F2F, NT	83.86
3D ResNet (with SE) [31]	FS, DF, F2F, NT	80.0
Timesformer [31]	FS, DF, F2F, NT	82.3
C2D	FS, DF, F2F, NT	84.2
CSN	FS, DF, F2F, NT	84.4
R(2+1)D	FS, DF, F2F, NT	84.6
V4D (modified)	**FS, DF, F2F, NT**	**84.6**

B. Single Manipulation Techniques

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:08 UTC from IEEE Xplore. Restrictions apply.

TABLE II. IDENTIFYING EACH MANIPULATION TECHNIQUE SEPARATELY, LQ. REPORTSON DF, F2F, FS, NT.

Algorithm	DF	F2F	FS	NT
Steg. Features + SVM [25]	73.64	73.72	68.93	63.33
Cozzolino et al. [26]	85.45	67.88	73.79	78.00
Bayar and Stamm [27]	84.55	73.72	82.52	70.67
Rahmouniet al. [28]	85.45	64.23	56.31	60.07
MesoNet [29]	87.27	56.20	61.17	40.67
XceptionNet [30]	96.36	86.86	90.29	80.67
3D ResNet [31]	91.81	89.6	88.75	73.5
3D ResNet (with SE) [31]	81.70	77.00	75.90	66.25
Timesformer [31]	90.7	73.9	82.1	67.5
C2D	94.5	90	93	80.5
CSN	93.75	88	93.5	77.75
R(2+1)D	93.75	87.75	91.5	78.5
V4D (modified)	**95**	**88.25**	**93.5**	**80.75**

TABLE III. INDIVIDUAL IDENTIFICATION OF EACH MODIFICATION METHOD, LQ. AREA UNDER THE ROC CURVE REPORTED

Algorithm	DF	F2F	FS	NT
Sabir [32]	96.9	94.35	96.3	-
V4D (modified)	**97**	92.4	95.8	-

TABLE IV. INDIVIDUAL IDENTIFICATION OF EACH MODIFICATION METHOD, LQ. ACCURACY REPORTED

Algorithm	DF	F2F	FS	NT
Meso-4 [29]	89.1	83.2	-	-
Mesoinception-4 [29]	91.7	81.3	-	-
Rossler [16]	94	-	93	-
V4D (modified)	**95**	**88**	**93.5**	-

VII. CONCLUSION

In the research report, several SOTA systems were compared utilizing various manipulation strategies for deepfake detection. It evaluated the action recognition-based adaptation of the V4D approach. Then, the strategy was expanded to include the detection of DeepFakes after this dataset was used to pretrain the algorithms. The experiments produced excellent results that demonstrated the superiority of the image-based forgery detection methods. Networks typically aren't good at transferring information from some of the manipulation method which are trained to the manipulation method which are tested, which is the cause of this. As a result, machine learning models may behave in an unforeseen manner.

REFERENCES

[1] "The Growing Threat of Deepfake Videos - SecurityWeek." https://www.securityweek.com/growing-threat-deepfake-videos/ (accessed Mar. 19, 2023).

[2] A. De Ruiter and A. D. Nl, "The Distinct Wrong of Deepfakes porn · Social identity," 123AD, doi: 10.1007/s13347-021-00459-2.

[3] S. Lyu, "DEEPFAKE DETECTION: CURRENT CHALLENGES AND NEXT STEPS," *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2020, Accessed: Mar. 19, 2023. [Online].

[4] H.-W. Hsu and J.-J. Ding, "Deepfake Algorithm Using Multiple Noise Modalities with Two-Branch Prediction Network," in *Proceedings, APSIPA Annual Summit and Conference 2021*. Accessed: Mar. 19, 2023. [Online].

[5] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-based forensic method for copy-move attack detection and transformation recovery," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3 PART 2, pp. 1099–1110, Sep. 2011, doi: 10.1109/TIFS.2011.2129512.

[6] I. J. Goodfellow *et al.*, "Generative Adversarial Nets", Accessed: Mar. 20, 2023. [Online]. Available: http://www.github.com/goodfeli/adversarial

[7] T. Afouras and J. S. Chung, "Deep Audio-Visual Speech Recognition." Accessed: Mar. 20, 2023. [Online]. Available: https://arxiv.org/pdf/1809.02108.pdf

[8] J. B. Awotunde, R. G. Jimoh, A. L. Imoize, A. T. Abdulrazaq, C. T. Li, and C. C. Lee, "An Enhanced Deep Learning-Based DeepFake Video Detection and Classification System," *Electronics 2023, Vol. 12, Page 87*, vol. 12, no. 1, p. 87, Dec. 2022, doi: 10.3390/ELECTRONICS12010087.

[9] P. Korshunov and S. Marcel, "DeepFakes: a New Threat to Face Recognition? Assessment and Detection." Accessed: Mar. 20, 2023. [Online]. Available: https://arxiv.org/pdf/1812.08685.pdf

[10] D. Tran, H. Wang, L. Torresani, and M. Feiszli, "Video Classification with Channel-Separated Convolutional Networks".

[11] L. Deng, H. Suo, and D. Li, "Deepfake Video Detection Based on EfficientNet-V2 Network," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/3441549.

[12] T. T. Nguyen *et al.*, "Deep Learning for Deepfakes Creation and Detection: A Survey", Accessed: Mar. 20, 2023. [Online]. Available: https://app.dimensions.ai

[13] A. Raza, K. Munir, and M. Almutairi, "A Novel Deep Learning Approach for Deepfake Image Detection," *Applied Sciences 2022, Vol. 12, Page 9820*, vol. 12, no. 19, p. 9820, Sep. 2022, doi: 10.3390/APP12199820.

[14] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[15] "FaceForensics/README.md at master · ondyari/FaceForensics · GitHub." https://github.com/ondyari/FaceForensics/blob/master/dataset/README.md (accessed Mar. 19, 2023).

[16] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images".

[17] S. Yadav, S. Bommareddy and D. K. Vishwakarma, "Robust and Generalized DeepFake Detection," 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2022, pp. 1-6, doi: 10.1109/ICCCNT54827.2022.9984553.

[18] L. Wang, Y. Xiong, Z. Wang, and Y. Qiao, "Towards Good Practices for Very Deep Two-Stream ConvNets," 2015, Accessed: Mar. 20, 2023. [Online]. Available: https://github.com/yjxiong/caffe/tree/action_recog

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:08 UTC from IEEE Xplore. Restrictions apply.

[19] K. Simonyan and A. Zisserman, "Two-Stream Convolutional Networks for Action Recognition in Videos".

[20] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A Closer Look at Spatiotemporal Convolutions for Action Recognition." Accessed: Mar. 20, 2023. [Online]. Available: https://arxiv.org/pdf/1711.11248.pdf

[21] D. Tran, H. Wang, L. Torresani, J. Ray, Y. Lecun, and M. Paluri, "A Closer Look at Spatiotemporal Convolutions for Action Recognition".

[22] D. Tran, H. Wang, L. Torresani, J. Ray, Y. Lecun, and M. Paluri, "A Closer Look at Spatiotemporal Convolutions for Action Recognition".

[23] Y. L. Chang *et al.*, "Consolidated Convolutional Neural Network for Hyperspectral Image Classification," *Remote Sensing 2022, Vol. 14, Page 1571*, vol. 14, no. 7, p. 1571, Mar. 2022, doi: 10.3390/RS14071571.

[24] C. Feichtenhofer, H. Fan, J. Malik, and K. He, "SlowFast Networks for Video Recognition", Accessed: Mar. 20, 2023. [Online]. Available: http://activity-net.org/

[25] J. Fridrich and J. Kodovský, "Rich Models for Steganalysis of Digital Images," 2012.

[26] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting Residual-based Local Descriptors as Convolutional Neural Networks," in *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security*, Jun. 2017, pp. 159–164. doi: 10.1145/3082031.3083247

[27] B. Bayar and M. C. Stamm, "A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer," in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, Jun. 2016, pp. 5–10. doi: 10.1145/2909827.2930786.

[28] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen, "Distinguishing computer graphics from natural images using convolution neural networks," in *2017 IEEE Workshop on Information Forensics and Security (WIFS)*, Dec. 2017, pp. 1–6. doi: 10.1109/WIFS.2017.8267647.

[29] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," Sep. 2018, doi: 10.1109/WIFS.2018.8630761.

[30] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Oct. 2016, [Online]. Available: http://arxiv.org/abs/1610.02357

[31] A. Das, S. Das, and A. Dantcheva, "Demystifying Attention Mechanisms for Deepfake Detection," in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, Dec. 2021, pp. 1–7. doi: 10.1109/FG52635.2021.9667026.

[32] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," May 2019.

[33] S. Dahiya, A. Gosain, and S. Mann, "Experimental analysis of fuzzy clustering algorithms," in *Advances in Intelligent Systems and Computing*, Singapore: Springer Singapore, 2021, pp. 311–320.

[34] S. Dahiya, "Using clustering techniques and classification mechanisms for fault diagnosis," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 2138–2146, 2020.

[35] S. Dahiya, R. Tyagi, and N. Gaba, "Comparison of ML classifiers for Image Data," EasyChair, 2020.

[36] A. Gosain and S. Dahiya, "An effective fuzzy clustering algorithm with outlier identification feature," *J. Intell. Fuzzy Syst.*, vol. 41, no. 1, pp. 2417–2428, 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:08 UTC from IEEE Xplore. Restrictions apply.

2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)

Investigating the Impact of Visual Attention Models in Face Forgery Detection

1st Ankit Yadav
Department of Information Technology
Delhi Technological University
Delhi, India
ankit4607@gmail.com

2nd Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological University
Delhi, India
dvishwakarma@gmail.com

2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC) | 979-8-3503-2379-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAISC58445.2023.10199338

***Abstract*— With the recent rise of realistic face manipulation methods, building robust face tampering detection methods has become more important than ever before. Visual attention has played an important role in highlighting discriminative regions within input which is important for making accurate predictions. This manuscript presents a comparative study of several recently proposed visual attention models for the problem of face forgery detection. Specifically, five visual attention models namely, coordinate, selective kernel, triplet, CoT, and shuffle attention have been tested by integrating with a baseline deep learning model. The modified visually attentive architectures are trained and tested on popular public benchmark dataset FaceForensics++. The experimental results achieved by different attention approaches are compared. Additionally, the computational costs involved in each type of attention have also been discussed specifying the accuracy and computation tradeoff. Experimental results prove that Triplet Attention performs best by achieving accuracy scores of 0.9543 and 0.7190 on DF and NT categories of the FF++ dataset. Triplet attention is also extremely lightweight with only 1200 trainable parameters compared to the other attention modules under study.**

Keywords*— *Visual attention, Face forgery, Face tampering, deepfake, attention, detection

I. Introduction

The emergence of realistic multimedia editing tools, in tandem with the emergence of big data from social media channels, has provided a perfect atmosphere for easy information tampering. Such tampering has the potential to mislead and disrupt society. The threat of changing identities in photographs and films, sometimes known as facial manipulation, is even greater. Realistic facial manipulation allows an innocent man's reputation to be harmed by making him appear guilty of illicit crimes. Several realistic face-altering technologies have recently been created.

Deepfakes [1] maps one person's face onto another via neural networks. FaceSwap [2] does the same thing through computer graphics. Face2Face [3] conveys a person's facial expression from one person to another. DeepFaceLab [4], a prominent combined library of similar tampering methods, includes a long number of identity tampering methods named face replacement, lip movement manipulation, de-aging, and so on. GANs (Generative Adversarial Networks) [5] have enabled the creation of totally artificial faces that do not persist in real world. Recent GAN techniques, such as ProGAN [6], StyleGAN [7], and others, can create hyper-realistic facial images with resolutions up to 1024×1024.

Various contributions have been proposed for detection of facial tampering. These methods extract a variety of distinct domain information such as spatial [8, 9, 10], textural [11, 12, 13], optical flow etc to make their predictions.

A. Contribution of this Manuscript

The major contributions of this manuscript are:

- Conduct an exhaustive study of recent visual attention models and demonstrate their effectiveness in detecting face forgery.
- Five attention models namely, Coordinate Attention [14], Selective Kernel Attention [15], Triplet Attention [16], CoT Attention [17] and Shuffle Attention [18] have been studied.
- The attention modules are evaluated by measuring their performance on the popular face forgery dataset FaceForensics++.
- A study of the computational complexity of each type of visual attention has been conducted.

B. Organization of this Manuscript

The rest of the manuscript is arranged as follows. Section II describes related work highlighting the existing literature on face tampering detection. Next, Section III presents several recently proposed visual attention methods. Section IV describes the experimental setup, including datasets, preprocessing, hardware, classification metrics, hyperparameter adjustment, and an approach for model weight initialization. Section V discusses the results obtained and presents a comparison of various attention models with regards to performance and computational overheads. Finally, Section VI specifies conclusion.

II. Related Works

This section discusses the available research for detecting face manipulation.

Guo et al. [19] proposed to extract better discriminative features by highlighting manipulation traces using gradient operators for each feature channel. Ke et al. [20] improve the detection capability of face tampering in low-quality degraded images by proposing a novel feature transformation module that upgrades features extracted from degraded images and transforms them to high-quality level. Yu et al. [21] propose to highlight multi-scale spatiotemporal clues both locally and globally. Caldelli et al. [22] reach 0.9100 accuracy on the FF++ dataset by using bi-directional optical flow mapping. To extract multi-scale features, Lin et al. [23] employ depthwise separable and dilated convolution kernels. To highlight forging indicators hidden in shallow characteristics, Zhao et al. [24] create a "textural enhancement block" that focuses on high-frequency components. Yang et al. [13] suggested a

979-8-3503-2379-5/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:17 UTC from IEEE Xplore. Restrictions apply.

unique architecture for detecting facial manipulation that extracts multi-scale textural data. Amerini et al. [25] employed the use of optical.

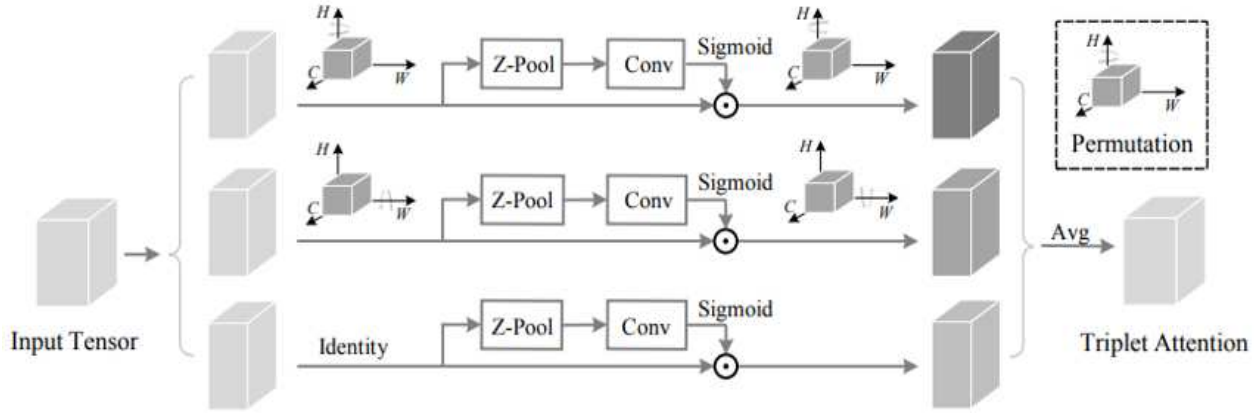


Fig. 1. Structure of Triplet Attention

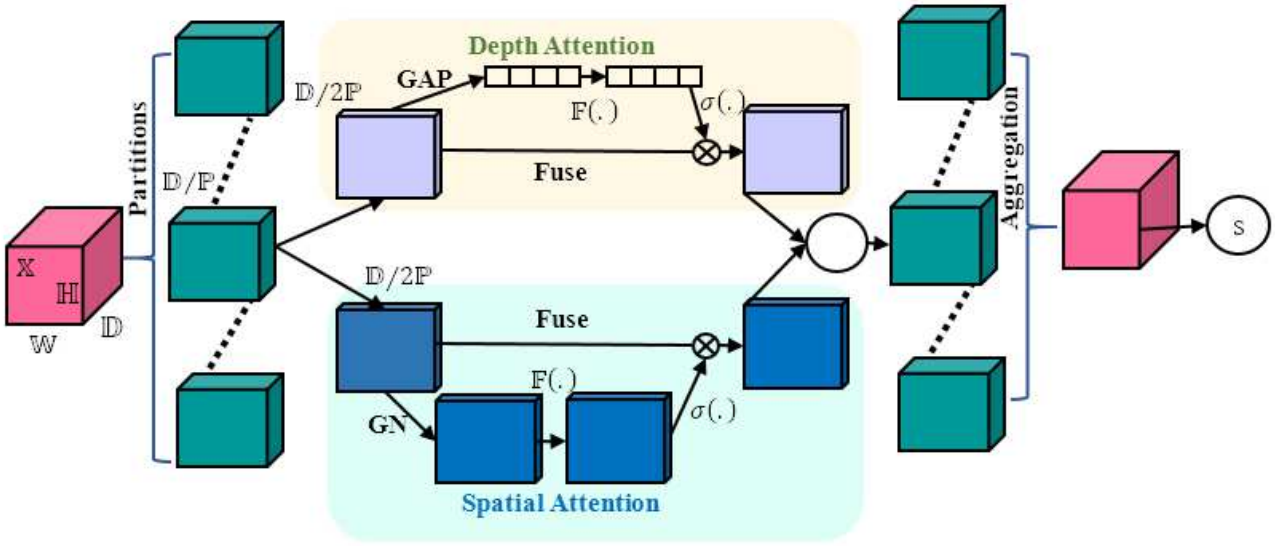


Fig. 2. Structure of Shuffle Attention

To detect manipulation, Guo et al. [26] extract a manipulation trace by suppressing the semantic content of images. Hu et al. [27] offer a lightweight video modification detection approach that takes only the I-frames of input videos into account. To detect manipulation in facial videos, Xu et al. [9] aggregate multi-frame characteristics. Lu et al. [28] provide a unique inception attentional module that incorporates channel and spatiotemporal feature learning. Ma et al. [29] use face video frames to extract lip characteristics and highlight temporal discrepancies. To detect manipulation, Nirkin et al. [30] emphasize the inconsistency of the facial region from surrounding areas such as hair, ears, neck, and so on in altered samples. Numerous GANs have also been utilized to generate completely synthetic images of faces [6, 7, 31].

III. Methodology

This section describes several attention modules and the baseline model used to evaluate them.

A. Shuffle Attention

Shuffle attention [18] proposes an efficient way to fuse attention mechanisms across the spatial and channel dimensions of features without increasing the computational cost. Specifically, a given input $\mathcal{X} \in \mathbb{R}^{H\times W\times C}$, is sub-divided into $\mathcal{G}$ sub-groups over channel dimension having a shape $\mathcal{X}_k \in \mathbb{R}^{H\times W\times C/\mathcal{G}}$. Each $\mathcal{X}_k$ is further sub-divided into 2-branches having shape $H \times W \times C/2\mathcal{G}$. The two branches compute spatial as well as channel attention respectively and the reduced computation is enforced by the subdivision of input along the channel dimension as shown in Fig. 2.

B. Selective Kernel (SK) Attention

Selective Kernel Attention [15] is a three-stage method that uses multiple-sized convolutional kernels for processing inputs with adaptive receptive fields. Specifically, it uses 3×3 and 5×5 kernels to extract multi-scale features in the 'split' stage. For input $\mathcal{X} \in \mathbb{R}^{H'\times W'\times C'}$ the method uses two convolutional kernels to produce $(\tilde{\mathcal{U}} \in \mathbb{R}^{H\times W\times C})$ and $(\hat{\mathcal{U}} \in \mathbb{R}^{H\times W\times C})$ feature maps. The 'fuse' stage fuses discriminative information from each receptive field branch using the summation operation and followed by global average pooling $\mathcal{GP}(.)$ along the channel dimension as given by Eq. 1:

$$\mathcal{U} = \mathcal{GP}(\tilde{\mathcal{U}} + \hat{\mathcal{U}}) \quad (1)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:17 UTC from IEEE Xplore. Restrictions apply.

Finally, the 'select' operation employs the soft attention mechanism across feature channels to highlight important channels.

C. CoT Attention

CoT attention [17] or Contextual Transformer attention improves the transformer-based self-attention mechanism by searching for important modeling dependencies among neighboring keys. Specifically, the proposed attention model combines self-attention learning with contextual feature mining by finding relevant relationships among neighboring keys for each query-key pair.

For input $\mathcal{X} \in \mathbb{R}^{H\times W\times C}$ and keys ($\mathcal{K} = \mathcal{X}$), query ($\mathcal{Q} = \mathcal{X}$) and value ($\mathcal{V} = \mathcal{X}\mathcal{W}_v$), the CoT block uses $\hbar \times \hbar$ grouped convolution to learn contextual keys ($\mathcal{K}^1 \in \mathbb{R}^{H\times W\times C}$) representing the relationship of neighboring keys. Finally, the attention matrix is derived by Eq 2.:

$$\mathcal{A} = [\mathcal{K}^1, \mathcal{Q}]\mathcal{W}_\theta\mathcal{W}_\delta \tag{2}$$

The final attentional feature maps are given by ($\mathcal{K}^2$) by Eq. 3:

$$\mathcal{K}^2 = \mathcal{V} \circledast \mathcal{A} \tag{3}$$

D. Triplet Attention

Triplet attention [16] is a three-branch structure that uses the 'rotation' operation to capture discriminative regions within input in three directions while keeping minimal computation overheads. Fig. 1 represents structure triplet attention module structure.

The triplet attention contains three branches. Two branches are responsible for processing the input of shape $\mathcal{X} \in \mathbb{R}^{H\times W\times C}$ and using a novel Z-pool layer to mine important channel features along the H and W dimension while the third branch works like a traditional spatial attention module.

The highlight of the triplet attention module is that it is extremely lightweight and its computation does not increase with increased input dimensions.

E. Coordinate Attention

Coordinate Attention [14] is a novel channel attention mechanism that captures long-range feature dependencies over one spatial dimension and also preserves exact positional encodings through another spatial dimension. Unlike common channel attention mechanisms that use 2D global pooling, coordinate attention uses two 1D kernels. For an input $\mathcal{X} \in \mathbb{R}^{H\times W\times C}$, two 1D pooling kernels of shape $\mathbb{R}^{H\times 1}$ and $\mathbb{R}^{1\times W}$ are used to encode discriminative features along the height and width in each channel thereby yielding direction-aware feature maps.

F. Baseline Architecture

To perform an effective comparison of the abovementioned attention mechanisms the baseline architecture from [32] has been used. Attention layers are attached after each bottleneck layer of this architecture.

IV. Experimental Setup

This section explains experimental steps taken to establish proposed model validity.

A. Datasets

This section describes the state-of-the-art publicly available datasets.

1) FaceForensics++ : The FaceForensics++ (FF++) [33] contains multiple face tampering examples such as Deepfakes (DF) [1] , FaceSwap (FS) [2] , Face2Face (F2F) [3], FaceShifter [34] and Neural Textures (NT) [35]. Each manipulation category contains 1000 videos created from 1000 original samples. Videos are provided in three qualities: raw, high (c23) and low (c40) compression. The c23 samples are utilized in this experiment.

In this paper, experiments are directed on DF and NT categories of FF++ dataset.

B. Classification Metrics

The classification metrics utilized in this experiment are given in the table below.

TABLE I. Classification Metrics Used in This Experiment

Metric	Formula	Range
Accuracy (ACC)	$\frac{\mathbb{TP}+\mathbb{TN}}{\mathbb{TP}+\mathbb{TN}+\mathbb{FP}+\mathbb{FN}}$	[0,1]
Precision (P)	$\frac{\mathbb{TP}}{\mathbb{TP}+\mathbb{FP}}$	[0,1]
Recall (R)	$\frac{\mathbb{TP}}{\mathbb{TP}+\mathbb{FN}}$	[0,1]
F1 score (F1)	$\boldsymbol{2 * \frac{Precision * Recall}{Precision + Recall}}$	[0,1]
Area Under Curve (AUC)	-	[0,1]
Mathews Correlation Coefficient (MCC)	$\frac{\mathbb{TP}*\mathbb{TN}-\mathbb{FP}*\mathbb{FN}}{\sqrt{(\mathbb{TP}+\mathbb{FP})(\mathbb{TP}+\mathbb{FN})(\mathbb{TN}+\mathbb{FP})(\mathbb{TN}+\mathbb{FN})}}$	[-1,1]

C. Hardware Specifications

Experiments are executed on two 24 GB NVIDIA A5000 GPUs in parallel. The system memory is 128GB.

D. Preprocessing

This section describes the preprocessing steps followed in this experiment.

1) Face Extraction: Popular deepfake detection algorithms have mostly used the dlib library [11, 9, 36] or MTCNN [37, 28, 38] for face detection and extraction. RetinaFace [34] is utilized in this experiment to extract facial images from video frames given its low failure rate compared to MTCNN [39].

2) Resizing, Normalization and Data Augmentation: Facial images that are cropped from video frames are resized to 128×128. Pixel values are normalized to range $[0,1]$. Facial images are flipped randomly in both vertical and horizontal directions with a flipping probability of 0.5.

E. Hyperparameters and Training Conditions

All experiments are run for 30 epochs. The batch size is set to 4. The Adam optimizer is used to update model weights. The initial learning rate is set to 0.01. A linear learning rate decay is employed that reduces learning rate by 10% after every 2 epochs.

F. Model Weight Initialization

Deep models when initialized with pre-trained ImageNet weights have performed better on classification tasks as compared to random weight initialization. Hence, the model weights in this experiment are initialized with ImageNet pre-trained weights provided on a Pytorch framework website.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:17 UTC from IEEE Xplore. Restrictions apply.

V. Results & Discussion

This section demonstrates the results obtained by each attention mechanism. Additionally, it also explains the computational cost involved for each attention mechanism.

A. System & Software Requirements

The proposed model requires at least 16GB memory of NVIDIA graphic card. The secondary storage required for storage of the FF++ dataset is 10GB. All coding experiments are conducted in the Python language. PyTorch framework is utilized to design and train neural network. Jupyter Notebook IDE has been used to write the code.

B. Results & Discussion

Table II represents performance of each attention mechanism on the DF (FF++) dataset. The triplet attention mechanism performs best with 0.9543 accuracy and 0.9874 AUC score. The shuffle attention achieves the highest Recall score and Coordinate attention is best in terms of the F1 score.

TABLE II. Results of each attention module on the DF (FF++) dataset. The best scores are highlighted in bold.

Attention Modules	ACC	P	R	F1	AUC	MCC
Coordinate	0.9408	0.9167	0.9589	0.9373	0.9842	0.8822
Selective Kernel	0.8901	0.9263	0.8532	0.8882	0.9649	0.7831
Triplet	0.9543	0.9415	0.9566	0.9490	0.9874	0.9077
CoT	0.9214	0.9224	0.9015	0.9118	0.9717	0.8411
Shuffle	0.8480	0.7656	0.9798	0.8596	0.9560	0.7248

Table III presents performance of the various attention mechanisms on the NT (FF++) dataset. Here again, Triplet attention has the best scores against other attention mechanisms. Coordinate attention gets the highest precision score of 0.7511 while triplet attention scores highest on all other metrics.

TABLE III. Results of each attention module on the NT (FF++) dataset. The best scores are highlighted in bold.

Attention Modules	ACC	P	R	F1	AUC	MCC
Coordinate	0.7104	0.7511	0.6763	0.7112	0.7917	0.4248
Selective Kernel	0.6479	0.6815	0.6269	0.6531	0.7008	0.2980
Triplet	0.7190	0.6486	0.9347	0.7658	0.8592	0.4897
CoT	0.6409	0.5952	0.8579	0.7028	0.7548	0.3165
Shuffle	0.6556	0.6149	0.7747	0.6856	0.7093	0.3262

The NT dataset is more challenging and hence the scores achieved are lower than those achieved on the DF dataset.

Fig. 3 and Fig. 4 present comparison of accuracy scores of various attention mechanisms on the DF and NT datasets of FF++ respectively. In both cases, Triplet Attention outperforms all other attention mechanisms while coordinate attention comes second in terms of performance.

Fig. 5 and Fig. 6 demonstrates AUC-ROC curve for attention mechanisms on the DF and NT datasets of FF++ respectively.

The Triplet attention achieves the highest AUC score of 0.9874 on DF dataset and 0.8592 on NT dataset, clearly stating that model is highly confident in its predictions.

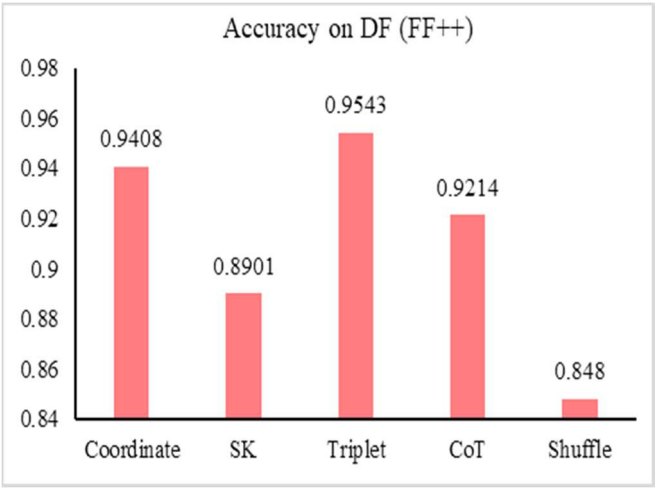


Fig. 3. Accuracy comparison of attention modules on the DF (FF++) dataset.

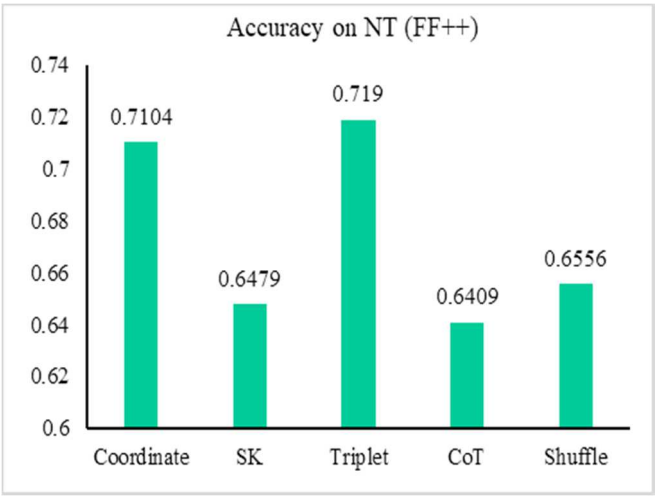


Fig. 4. Accuracy comparison of attention modules on the NT (FF++) dataset

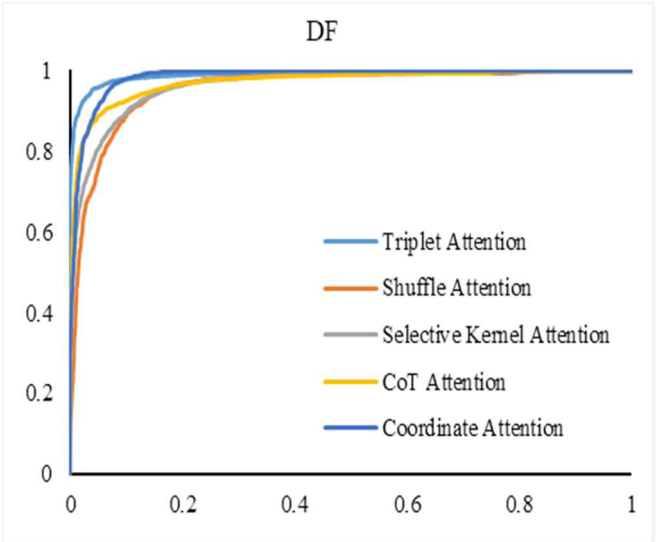


Fig. 5. AUC curves for various attention modules on the DF (FF++) dataset.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:17 UTC from IEEE Xplore. Restrictions apply.

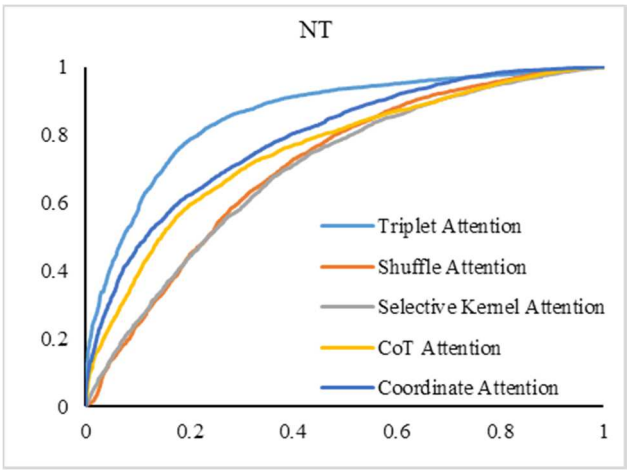


Fig. 6. AUC curves for various attention modules on the NT (FF++) dataset

C. *Complexity Analysis*

This section presents a discussion of the computational overheads involved in implementing each attention mechanism. As discussed earlier, each attention mechanism has been integrated by using four attention layers, one for each skip connection block of the baseline model.

Fig. 8 presents a comparison of number of parameters for each type of attention mechanism. The baseline model without any attention layers has 23.5 million parameters (blue bar). Triplet, Coordinate and Shuffle attention layers (green bars) are extremely lightweight having only 530280, 1200 and 1440 parameters respectively. Selective Kernel (SK) and CoT attention (green bars) are computationally heavy having 61.3 million and 48.7 million parameters respectively.

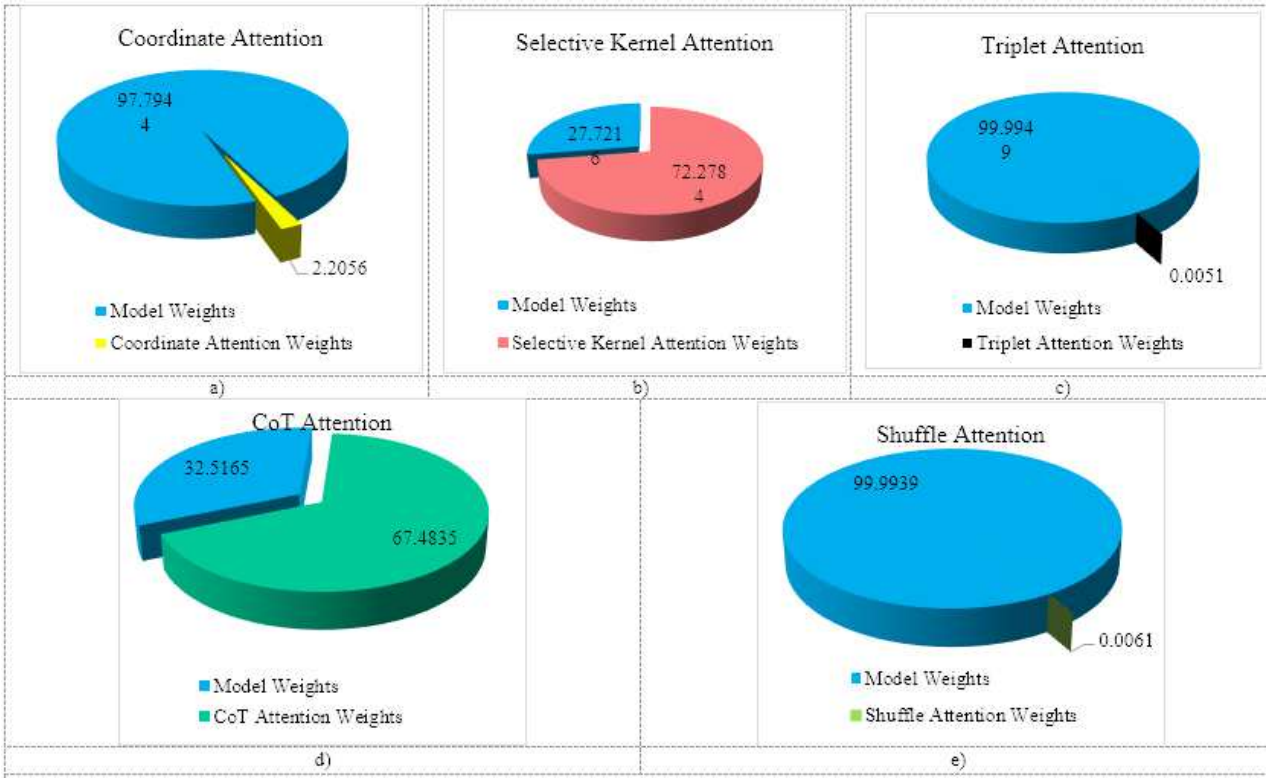


Fig. 7. The parameter distribution of each attention mechanism. The values represent the percentage of model parameters used for attention mechanisms. Coordinate Attention (a), Triplet Attention (c) and Shuffle Attention (e) are extremely lightweight occupying only 2.2056%, 0.0051% and 0.0061% of total model parameters respectively. Selective Kernel attention (b) and CoT attention (d) are heavier attention modules having 72.2784% and 67.4835% of the total model parameters.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:17 UTC from IEEE Xplore. Restrictions apply.

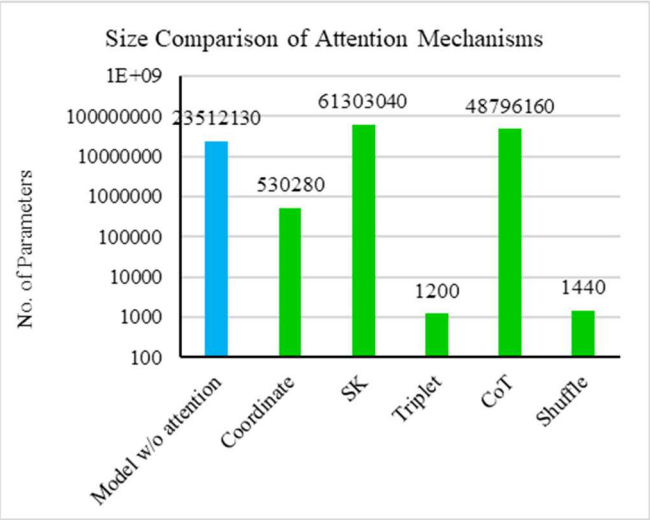


Fig. 8. Size comparison of attention mechanisms in terms of th number of parameters.

Fig . 7 represents the percentage distribution of parameters for each type of attention mechanism. The Coordinate, Triplet and Shuffle attention are lightweight occupying only 2.2056%, 0.0051% and 0.0061% of total model parameters respectively. The Selective Kernel attention and CoT attention are heavier attention modules having 72.2784% and 67.4835% of the total model parameters.

VI. Conclusion

This manuscript presents a comparison of the recently proposed visual attention models for face forgery detection. Specifically, a baseline CNN has been enhanced with several recently proposed attention modules and their impact has been studied. The study has been conducted in terms of the performance on the FaceForensics++ dataset. Additionally, the computational complexity of different attention mechanisms has also been analyzed and compared visually. It is clear from the conducted experiment that different attention modules employ different computational overheads to existing CNN parameters to highlight important spatial regions of input.

A. Limitations

This paper used a baseline CNN and enhances it using different types of visual attention modules. There are several recently proposed visual attention mechanisms that are not included in this experiment. More recent visual attention modules can improve the model performance further. Other vision baseline models can be integrated with vision based attention approaches for boosting model performance even further.

B. Future Directions

While vision based attention mechanisms have improved deepfake detection, multi-modal approaches are a promising area to enhance the forgery detection process. Specifically, researchers can focus on other modalities such as text or audio that can further improve the tampering detection capablities of models. Attention mechanism can be used to combine multi-modal feature extraction for better discriminative capabilities.

References

[1] ["DeepFakes," GitHub, 14 August 2020. [Online]. Available: https://github.com/deepfakes/faceswap. [Accessed 08 July 2022].

[2] "FaceSwap," GitHub, 19 June 2016. [Online]. Available: https://github.com/MarekKowalski/FaceSwap. [Accessed 08 July 2022].

[3] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt and M. Nießner, "Face2Face: real-time face capture and reenactment of RGB videos," Communications of the ACM, vol. 62, no. 1, pp. 96-104, 2019.

[4] "DeepFaceLab," GitHub, 18 March 2020. [Online]. Available: https://github.com/iperov/DeepFaceLab. [Accessed 08 July 2022].

[5] I. Goodfellow, J. P.-. Abadie, M. Mirza, B. Xu, D. W.-. Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Nets," in Advances in Neural Information Processing Systems (NIPS 2014), Montreal, 2014.

[6] T. Karras, T. Aila, S. Laine and J. Lehtinen, "Progressive Growing of GANs for Improved Quality, Stability, and Variation," in International Conference on Learning Representations, Vancouver Convention Center, Vancouver, BC, Canada, 2018.

[7] T. Karras, S. Laine and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019.

[8] Z. Shang, H. Xie, Z. Zha, L. Yu, Y. Li and Y. Zhang, "PRRNet: Pixel-Region relation network for face forgery detection," Pattern Recognition, vol. 116, 2021.

[9] Z. Xu, J. Liu, W. Lu, B. Xu, X. Zhao, B. Li and J. Huang, "Detecting facial manipulated videos based on set convolutional neural networks," Journal of Visual Communication and Image Representation, vol. 77, 2021.

[10] Z. Chen and H. Yang, "Attentive Semantic Exploring for Manipulated Face Detection," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, 2021.

[11] M. Bonomi, C. Pasquini and G. Boato, "Dynamic texture analysis for detecting fake faces in video sequences," Journal of Visual Communication and Image Representation, vol. 79, 2021.

[12] J. Yang, S. Xiao, A. Li, G. Lan and H. Wang, "Detecting fake images by identifying potential texture difference," Future Generation Computer Systems, vol. 125, pp. 127-135, 2021.

[13] J. Yang, A. Li, S. Xiao, W. Lu and X. Gao, "MTD-Net: Learning to Detect Deepfakes Images by Multi-Scale Texture Difference," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 4234 - 4245, 2021.

[14] Q. Hou, D. Zhou and J. Feng, "Coordinate Attention for Efficient Mobile Network Design," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021.

[15] X. Li, W. Wang, X. Hu and J. Yang, "Selective Kernels Network," in Conference on Computer Vision and Pattern Recognition, Long Beach, California, 2019.

[16] D. Misra, T. Nalamada, A. U. Arasanipalai and Q. Hou, "Rotate to Attend: Convolutional Triplet Attention Module," in IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2021.

[17] Y. Li, T. Yao, Y. Pan and T. Mei, "Contextual Transformer Networks for Visual Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 2, pp. 1489-1500, 2023.

[18] Q.-L. Zhang and Y.-B. Yang, "SA-Net: Shuffle Attention for Deep Convolutional Neural Networks," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021.

[19] Z. Guo, G. Yang, D. Zhang and M. Xia, "Rethinking gradient operator for exposing AI-enabled face forgeries," Expert Systems with Applications, vol. 215, 2023.

[20] J. Ke and L. Wang, "DF-UDetector: An effective method towards robust deepfake detection via feature restoration," Neural Networks, vol. 160, pp. 216-226, 2023.

[21] Y. Yu, X. Zhao, R. Ni, S. Yang, Y. Zhao and A. C. Kot, "Augmented Multi-Scale Spatiotemporal Inconsistency Magnifier for Generalized DeepFake Detection," IEEE Transactions on Multimedia, vol. Early Access, pp. 1-13, 2023.

[22] R. Caldelli, L. Galteri, I. Amerini and A. D. Bimbo, "Optical Flow based CNN for detection of unlearnt deepfake manipulations," Pattern Recognition Letters, vol. 146, pp. 31-37, 2021.

[23] H. Lin, W. Huang, W. Luo and W. Lu, "DeepFake detection with multi-scale convolution and vision transformer," Digital Signal Processing, vol. 134, 2023.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:17 UTC from IEEE Xplore. Restrictions apply.

[24] H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen and N. Yu, "Multi-attentional Deepfake Detection," in Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021.

[25] I. Amerini, L. Galteri, R. Caldelli and A. D. Bimbo, "Deepfake Video Detection through Optical Flow Based CNN," in IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, 2019.

[26] Z. Guo, G. Yang, J. Chen and X. Sun, "Fake face detection via adaptive manipulation traces extraction network," Computer Vision and Image Understanding, vol. 204, 2021.

[27] J. Hu, X. Liao, W. Wang and Z. Qin, "Detecting Compressed Deepfake Videos in Social Networks Using Frame-Temporality Two-Stream Convolutional Network," IEEE Transactions on Circuits and Systems for Video Technology , 2021.

[28] C. Lu, B. Liu, W. Zhou, Q. Chu and N. Yu, "Deepfake Video Detection Using 3D-Attentional Inception Convolutional Neural Network," in IEEE International Conference on Image Processing (ICIP), Anchorage, 2021.

[29] J. Ma, S. Wang, A. Zhang and A. W.-C. Liew, "Feature Extraction For Visual Speaker Authentication Against Computer-Generated Video Attacks," in IEEE International Conference on Image Processing (ICIP), Abu Dhabi, 2020.

[30] Y. Nirkin, L. Wolf, Y. Keller and T. Hassner, "DeepFake Detection Based on Discrepancies Between Faces and their Context," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[31] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017.

[32] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016.

[33] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niessner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019.

[34] L. Li, J. Bao, H. Yang, D. Chen and F. Wen, "FaceShifter: Towards High Fidelity And Occlusion Aware Face Swapping," https://arxiv.org/abs/1912.13457, 2019.

[35] J. Thies, M. Zollhöfer and M. Nießner, "Deferred neural rendering: image synthesis using neural textures," ACM Transactions on Graphics, vol. 38, no. 4, pp. 1-12, 2019.

[36] J. Guo and Y. Liu, "Facial parts swapping with generative adversarial networks," Journal of Visual Communication and Image Representation, vol. 78, 2021.

[37] Y. Zhou, A. Luo, X. Kang and S. Lyu, "Face Forgery Detection Based On Segmentation Network," in IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 2021.

[38] J. Zhang, J. Ni and H. Xie, "DeepFake Videos Detection Using Self-Supervised Decoupling Network," in IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, 2021.

[39] J. Deng, J. Guo, E. Ververas, I. Kotsia and S. Zafeiriou, "RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:17 UTC from IEEE Xplore. Restrictions apply.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Isolated DC-DC Converter with Synchronous Rectification for Electric Vehicle Battery Charging

Awanish Akash
Department of Electrical Engineering
Delhi Technological University,
Delhi, India
akashawanish29@gmail.com

Mini Sreejeth
Department of Electrical Engineering
Delhi Technological University,
Delhi, India
minisreejeth@dce.ac.in

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205613

***Abstract*—The chargers for Electric Vehicles (EVs) plays a vital role in the development of the EV industries. EVs hold great promise for significantly decreasing the environmental impact of automobile transportation. The EV chargers have an important unit as a DC-DC converter. This paper proposes a closed-loop control approach for a PSFB (phase-shifted full bridge) dc-dc converter operating under synchronous rectification for EV charging applications. To achieve high efficiency, PSFB with synchronous Rectification is preferred over conventional PSFB. Zero voltage switching (ZVS) is supported by the converter, which enables it to operate at high frequencies while also providing benefits such as a solution to slow charging and high efficiency. The parasitic capacitances of full bridge MOSFETS and the leakage inductor of the high-frequency transformer play a key role in achieving ZVS. A 48V Lithium Ion Battery has been used for the charging process and one PI controller has been implemented for CC (Constant Current) mode charging. The system is designed for 5.5 kW, and the analysis is carried out using MATLAB/Simulink software.**

Keywords - dc-dc converter, EV battery charging, isolated transformer, switching frequency, synchronous rectification, ZVS

I. INTRODUCTION

The demand for Electric vehicles is supposed to increase in the upcoming years and it is forecasted that 10% of the total vehicle market will be dominated by EVs [1]. In comparison to traditional Internal Combustion (IC) engines, EVs are a better alternative as it is environmentally friendly. The use of Traditional IC engine vehicles emits lots of harmful gases like Carbon monoxide (CO) which pollutes the air, thereby playing a major role in Global warming [2]. For sustainable development, we should find a better alternative way that reduces environmental problems. The use of traditional IC engines has degraded air quality, which has a major impact on public health. The use of fossil fuels like petrol and diesel must be limited as it has limited availability. The extensive dependency on Fossil fuels has to be taken care of by switching to eco-friendly EVs. In terms of efficiency, 60 % of the total electrical energy available from the grid is converted to useful work of powering the wheel whereas, in the case of traditional IC engine vehicles, only 17% to 20% of total energy available from the fuel is used in powering the wheel which shows that there is 80% wastage of the fuel energy. The EVs have zero emissions which is good for the atmosphere. Even if the production of electricity is considered for EVs then IC engines have 3 times more carbon monoxide emissions than the average EVs. EVs have lower running costs and maintenance. By 2030, India is about to achieve 40% of its electric power from non-fossil fuel energy resources thereby, it will prove as a boon for EVs [3].

As the EV demand is increasing so there is a need for an electric charger that helps in charging the vehicle's battery There is a requirement for Fast charging stations so that the use of vehicles becomes more convenient. DC-DC converter is a basic unit of EV chargers playing a key role in charging. It is the connecting link between the AC/DC power transfer circuit and the EV battery. DC-DC converters help in attaining from one voltage level to another voltage level. For the conversion to occur, the energy is stored temporarily in energy-storing elements like inductors or capacitors, and this energy is released at different voltage levels to the output. In an inductor/high-frequency transformer the energy is stored in the form of magnetic energy and in a capacitor the energy is stored in the form of an electric field. There are different topologies available in DC-DC converters as per the requirement of voltage levels and efficiencies. Based on safety and galvanic isolation DC converters are classified as Isolated converters and non-Isolated converters. Isolated converters use high-frequency transformers and provide galvanic isolation. Converters like LLC, and DAB come under this category. Some examples of non-isolated converters are Buck-Boost converters, flyback converters, and CUK converters. The basic stages of charging are from ac (from the grid) to dc and then dc to dc conversion. Onboard EV chargers directly take ac from the supply and the later stages of conversion ac to dc and then dc to the required dc for the battery is present inside the vehicle charging system whereas, in Offboard chargers, DC is directly fed to the vehicle battery system, where DC-DC conversion occurs where DC-DC converters play a vital role in charging [4],[5]. The working of the converter also influences the lifetime of the battery.

Both ac and dc charging is possible; ac charging often specifies level 1 and level 2 onboard chargers whereas dc charging is specified by level 3 off-board chargers. Level 1 is simple charging available at residential, buildings where 230V ac is available. It supports 12 A to 15 A of current (Single Phase). Level 1 charging is slow charging. Level-2 ac chargers can produce up to 20kW of power and are typically utilized in commercial locations like malls and workplaces. Modern Level-2 charging stations accept 208Vac or 240Vac as input voltage. Level 3 chargers are fast offboarding DC-DC chargers [6],[7]. Figure 1 shows the stages of the charging system. It

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:06 UTC from IEEE Xplore. Restrictions apply.

depicts the power level conversion stage. The input to the system is dc and then DC is converted to AC which is fed to the high-frequency transformer of the PSFB converter later the output from the high-frequency transformer is converted to DC and this DC after passing through the LC filter is fed to the battery for charging purpose. Generally, Input DC is available from AC after rectification but here in this paper focus has been made on DC-DC conversion.

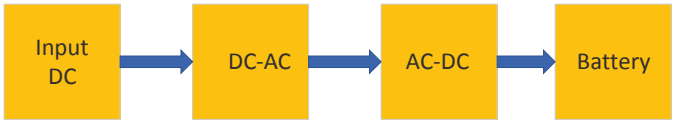


Fig. 1. Stages of conversion

PSFB with synchronous rectification is a good selection for EV chargers as it has the soft-switching capability, low stress on switches, low and EMI, simple PWM control. It has high efficiency and a wide range of ZVS than the conventional PSFB with diode rectification. It allows working at high frequency resulting in high power density [8]-[10]. The closed loop PI control has been used to achieve CC mode charging. The PI controller controls the pulses which is fed at the MOSFET switches.

Section II describes the model of the battery charging system along with filter design, PI controller, and analysis of ZVS. Section III covers a detailed analysis, and results. The study is validated using MATLAB/SIMULINK for a 5.5 kW system, 48V lithium-ion battery.

II. Design and Analysis of the Battery Charging System

The block diagram of the proposed charging consists of Input dc, dc-ac conversion, ac-dc conversion, Battery, and CC controller. Input DC is fed to the Phase shifted full bridge converter which is then converted to the EVs battery voltage level. AC power from the grid can be rectified to DC and fed at the input DC level. Other forms of renewable energy like solar energy can be used to feed power at input DC. In the Block diagram, different stages have been shown under the block Phase shifted Full bridge Converter. The first stage is DC to AC conversion with the help of the H bridge structure present in PSFB. This ac is fed to the high-frequency transformer which stores energy in the form of magnetic energy and builds the voltage level. In the last stage, this ac is converted to dc with the help of synchronous rectification where two MOSFETs are used at the secondary side of the high-frequency transformer. Here, Synchronous Rectification helps in attaining the dc voltage level at high efficiency. This dc is passed through the LC filter.

With the help of an LC filter, the output dc current and voltage smoothen, thereby, feeding better quality charging current to the battery. Figure 2 shows the Block diagram of the proposed Charging system. The CC controller consists of a PI controller which helps in maintaining the constant required current to the battery and also protects the battery from overheating.

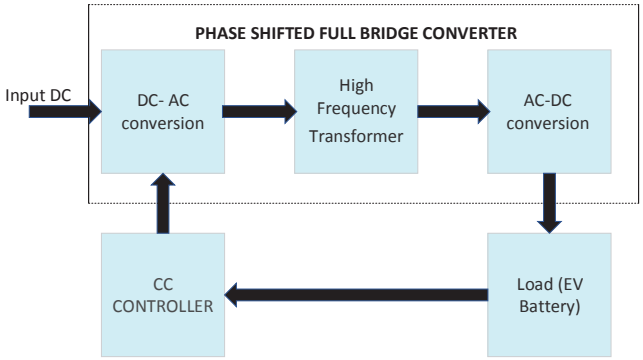


Fig. 2. Block diagram of Proposed Charging System

A. Phase – shifted full bridge DC-DC converter

The Phase Shifted Full bridge DC-DC Converter with synchronous Rectification is used for the proposed battery charging system. It has two stages of power conversion – Dc to ac(fed to a high-frequency transformer) and then ac to dc which is fed to the lithium-ion battery. Rather than driving both the diagonal switches together as in a Full bridge converter, here, a delay is introduced at the turn-on commands of the diagonal switches. The effective duty cycle can be controlled by modifying the phase shift between the switch drive commands. The Uniqueness of this phase shift technique is that two of switches connected in series with the transformer can be turned ON, yet the input voltage fed to the transformer is zero. These are either the two upper or two lower switches, not the diagonal switches, of the full bridge converter. The transformer primary is efficiently short-circuited and clamped to the appropriate input rail in this mode. The necessary off time for that specific switching cycle relates to the duration of time switches can be held in this condition. The primary current passes through the switch output capacitance (Coss) when the right switch is later switched off, causing the switch drain voltage to resonate with the opposing input rail.

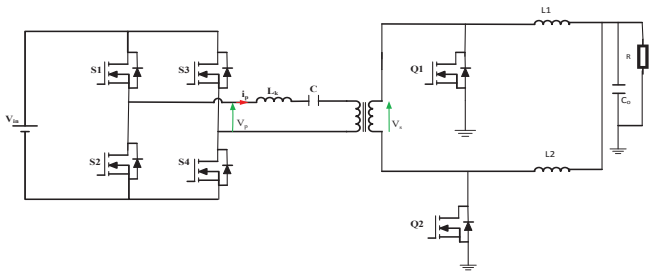


Fig. 3. PSFB with Synchronous Rectification

The PSFB with synchronous rectification as shown in Figure 3 has two MOSFET switches at the secondary side. Controlling the switch at the secondary side has a significant role. The MOSFET R_{dson} (drain to source resistance) can be adjusted, thereby, minimizing the conduction losses. The Converter allows ZVS resulting in minimum switching losses.

B. Filter design

The capacitor inductor filter at the output side is designed using the following equations. The voltage drop across the output inductor is given as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:06 UTC from IEEE Xplore. Restrictions apply.

$$V_L = V_{in}\frac{N_s}{N_p} - Vo \quad (1)$$

V_o is the output voltage

$V_{in}\frac{N_s}{N_p}$ is the secondary side transformer voltage with turns ratio N_s/N_p

$$V_L = \frac{di}{dt} \quad (2)$$

Using equations 1 and 2 , the inductance L can be written as

$$L = V_L\frac{DT_s}{I_{ripple}} \quad (3)$$

The output current through the filter capacitor is given as

$$I_{ripple} = C\frac{dv}{dt} \quad (4)$$

Equation 4 can be further written as

$$C = I_{ripple}\frac{DT_s}{V_{ripple}} \quad (5)$$

C. PI Controller Design and Operation

The closed-loop control of the battery charging current is done using a PI(also, known as Proportional Integral) controller. PI controller is a feedback control loop that helps in setting the desired value at the output. The reference desired value is set and with the help of the sensor, the output charging current of the battery is sensed. The difference between the reference value of the current and the output charging current known as the error signal is fed to the PI controller. This helps in controlling the switching pulses given to Switches, thereby controlling the operation of switches accordingly and helping to achieve a uniform charging current for the battery.

The mathematical equation of the PI controller is given by

$$u(t) = k_p e(t) + k_i\int_0^t e(t)dT \quad (6)$$

where, u(t) is the controlling signal, e(t) is the error signal

k_p is the proportional gain constant of P -term

k_i is the integral gain of I-term

The combination of proportional and integral signals is known as the controlling signal u(t). The setting of K_p and K_i is important for tuning the PI control and hence achieving the desired charging current at the battery side.

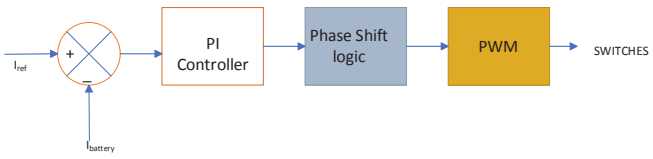


Fig. 4. Block diagram of PI control strategy

Figure 4 shows the block diagram of the PI control strategy. The diagram states the control of the charging current of the battery. PI controller helps in improving steady-state error characteristics for the system.

D. Analysis of ZVS and specifications

For achieving ZVS, the leakage inductance stored energy should be greater than the energy stored in output capacitances which is described in equation (7)

$$E = \frac{1}{2}L_{lk}I^2 > \frac{4}{3}C_{MOS}V_{in}^2 + \frac{1}{2}C_{tr}V_{in}^2 \quad (7)$$

where, C_{MOS} is the output capacitance of the switch

C_{tr} is the transformer winding capacitance

The equation above can be written in a simplified form:

$$\frac{1}{2}L_{lk}I^2 > \frac{1}{2}C_{eq}V_{in}^2 \quad (8)$$

On solving further, input primary current can be written as:

$$I = \sqrt{\frac{C_{eq} * V_{in}^2}{L_{lk}}} \quad (9)$$

The formula below gives the frequency required to achieve ZVS switching resonance between the transformer's leakage inductance, which stores energy, and the MOSFETs' junction capacitance.

$$fr = \frac{1}{2\pi\sqrt{L_{lk}C_r}} \quad (10)$$

TABLE I. LISTS THE SYSTEM PARAMETERS

Parameters	Units	Value
Input voltage	V	380
Output voltage	V	55
Switching frequency	kHz	200
Output current ripple	%	20
Output voltage ripple	%	40
Power Rating	kW	5.5
Battery Nominal Voltage	V	48
Battery Ah rating	Ah	100Ah
Battery Initial SOC	%	40

Depending on the requirement, batteries can be charged at different rates. Appropriate safety measures must be taken in order to maximize the charging rate, ensure the battery is fully charged, and prevent overcharging therefore, the system is operated under CC mode to charge 48V,100Ah battery

III. Simulation Results

The MATLAB Simulink environment is used to model the proposed charging system. The model consists of, a PSFB DC-DC converter, a PI controller, Lithium Ion Battery. The input DC voltage is 380 V, the output at the converter is 55 V, Switching Frequency is 200 kHz is used to charge a 48 V, 100 Ah Lithium-ion Battery

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:06 UTC from IEEE Xplore. Restrictions apply.

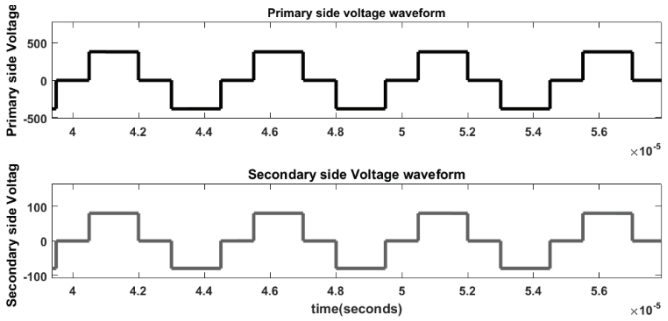


Fig. 5. Primary and secondary side voltage waveform

In Figure 5, the primary and secondary side voltage waveform has been shown. The PI controller feedback loop controls the switches in order to maintain CC mode and the phase shift between the diagonal switches has been maintained which helps in attaining the desired voltage waveform at the primary and secondary sides. On both sides, the transformer has zero voltage due to the two upper switches or two lower switches.

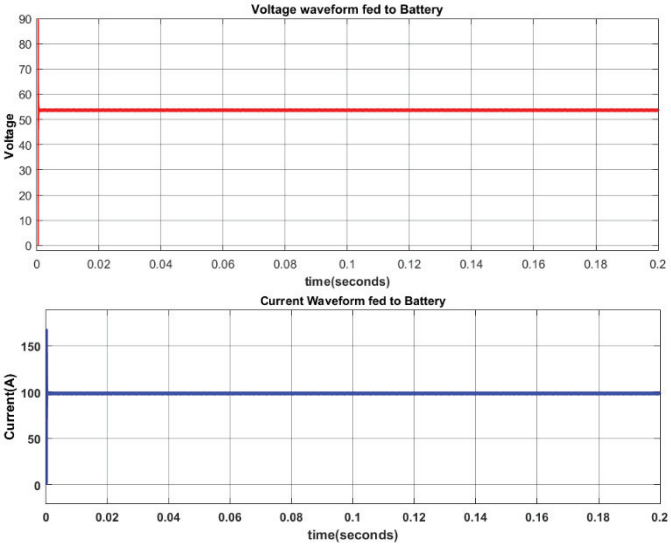


Fig. 6. Output Voltage and current waveform fed to Battery

In Figure 6, the output voltage and current waveform is shown which comes after synchronous rectification through the secondary side of the high-frequency transformer. The output voltage is 55 V at the output of the converter and it is fed to the battery. The lithium-ion battery charges with a charging current of 100 A at initial SOC of 40 %.

Fgure 7 shows the SOC, Current, and Voltage waveform when the SOC is moving from 40% to 50%. The magnitude of the charging current remains almost near 100 A. As the SOC increases the battery charges. When the battery is fully charged then SOC becomes constant. In the charging state, the battery shows a negative current.

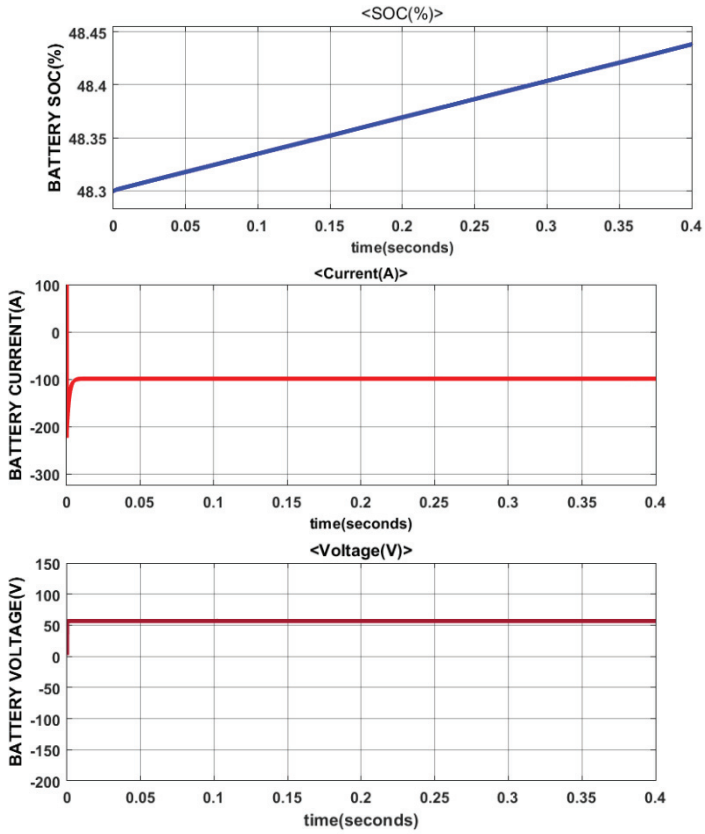


Fig. 7. SOC, Current, Voltage waveform from the Battery

The charging current remains almost constant up to 60% to 70 % of SOC and then the charging current starts decreasing gradually and at 99% SOC, the charging current becomes almost zero.

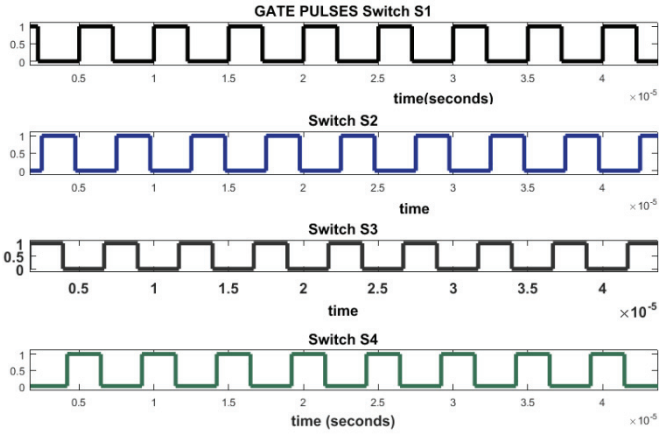


Fig. 8. Switching Pulses

Figure 8 shows the switching pulses for switches S_1, S_2, S_3, S_4. The closed loop PI control has controlled the switching pulses in order to maintain the desired output.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:06 UTC from IEEE Xplore. Restrictions apply.

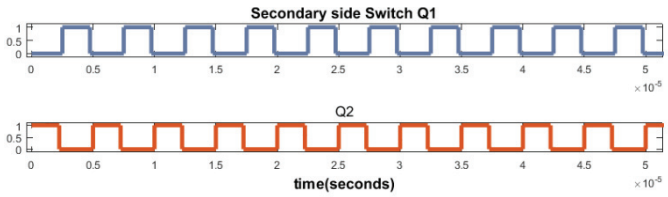


Fig. 9. Switching Pulse at Q1, Q2

Figure 9 shows the switching pulse waveform at the secondary side switches Q_1 and Q_2.

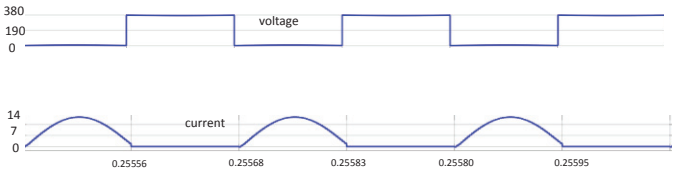


Fig. 10. Across Switch S_1 ZVS waveform

In Figure 10, Zero Voltage Switching has been shown across Switch S_1. During this duration when the voltage is zero across the switch then the switch gets turned ON resulting in minimum switching losses which help in efficient charging of the battery.

IV. CONCLUSION

In the proposed model, the charging of a 48V lithium-ion Battery through PSFB dc-dc converter closed-loop control has been studied in a SIMULINK environment. One PI controller has been used to maintain CC mode charging. The PI controller has a significant role in avoiding overshoots and undershoots for the charging current and has a faster response. SOC, current, and voltage waveforms of the battery have been presented for the charging state. PSFB dc-dc converter with synchronous rectification can be used for EV battery charging as it has high efficiency and enhances ZVS over a wide load range. The CC mode also helps in maintaining the life of the battery. The future scope of this paper can be extended to CC-CV charging and other Artificial Intelligence (AI) based control methods can be implemented.

REFERENCES

[1] Forecasting report of IHS automotive Q4/2018, 2018.

[2] B. Maclnnis and J. A. Krosnick. (Oct. 2020). Climate insights 2020: Electric vehicles. Resources for the Future (RFF). [Online]. Available: https://www.rff.org/publications/reports/climateinsights2020- electric-vehicles.

[3] e-amrit.niti.gov.in/benefits-of-electric-vehicles

[4] S. Harika, R. Seyezhai, and A. Jawahar, "*Investigation of DC Fast Charging Topologies for Electric Vehicle Charging Station (EVCS)*," IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON, vol. 2019-Octob, pp. 1148–1153, 2019, doi: 10.1109/TENCON.2019.8929621.

[5] Expo. - APEC, pp. 1127– 1133, 2012 [5] S. Chauhan and N. R. Tummuru, "*High frequency AC Link Based Isolated Dual Active Bridge DC-DC Converter Control, Features and Its Functionalities*," in Proc. Of IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020), Cochin, India, 2020, pp. 1-6.

[6] M. Safayatullah, M. T. Elrais, S. Ghosh, R. Rezaii and I. Batarseh, "*A Comprehensive Review of Power Converter Topologies and Control Methods for Electric Vehicle Fast Charging Applications*," in IEEE Access, vol. 10, pp. 40753-40793, 2022, doi: 10.1109/ACCESS.2022.3166935.

[7] Chakraborty, Sajib et al. "*DC-DC Converter Topologies for Electric Vehicles, Plug-in Hybrid Electric Vehicles and Fast Charging Stations: State of the Art and Future [1]Trends*." *Energies* (2019).

[8] S. Chothe, R. T. Ugale and A. Gambhir, "*Design and modeling of Phase Shifted Full Bridge DC-DC Converter with ZVS*," 2021 National Power Electronics Conference (NPEC), Bhubaneswar, India, 2021, pp. 01-06, doi: 10.1109/NPEC52100.2021.9672529.

[9] Texas Instruments; *Phase-Shifted Full-Bridge Controller with Synchronous Rectification* (2012).

[10] Ş. Küçük and E. Akboy, "*A Basic Phase Shift Full Bridge DC-DC Converter Design And Simulation*," 2022 57th International Universities Power Engineering Conference (UPEC), Istanbul, Turkey, 2022, pp. 1-7, doi: 10.1109/UPEC55022.2022.9917909.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:06 UTC from IEEE Xplore. Restrictions apply.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-16327-4

Legitimate and spam SMS classification employing novel Ensemble feature selection algorithm

Shailender Kumar[1] · **Shweta Gupta**[1]

Received: 7 March 2022 / Revised: 10 July 2023 / Accepted: 17 July 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
The classification of Short Message Service (SMS) is a crucial task that requires distinguishing undesirable and harmful messages from legitimate ones. These spam messages may include malicious links that reveal users' personal data to some sharp practices. Many Machine Learning (ML) and deep learning algorithms prevail for spam detection. However, a large number of features present in the message drop the classifier's performance, which can be untangled through feature engineering. This paper proposes an ensemble feature selection algorithm assessed over a high-dimensional imbalanced dataset that resembles the actual data prevailing in the surroundings. The proposed work employs wrapper-based Stratified K-fold Cross-Validated (SKCV) Recursive Feature Elimination with Cross-Validation (RFECV) over hyperparameters tuned Support Vector Classifier (SVC) and random forest algorithms. This study contributes an efficient feature selection algorithm that produces a pruned feature subset substituting the whole dataset. To establish the potential of the proposed approach, its performance is evaluated against five feature selection techniques, including Mutual Information (MI), Low Variance (LV), Particle Swarm Optimization (PSO), Chi-square, and Recursive Feature Elimination (RFE). All obtained feature subsets are analyzed with three ML models Logistic Regression (LR), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB). This research also compares the proposed framework results with Long Short Term Memory (LSTM) neural network to validate its competence.

Keywords Spam · Feature selection · Machine learning · Ensemble model · Optimal feature subset

✉ Shweta Gupta
Shwetag226@gmail.com

Shailender Kumar
Shailenderkumar@dce.ac.in

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

1 Introduction

With the emergence of telecommunication, mobile services have become a convenient amenity amongst users belonging to academics, corporations, or private grounds. As given in[1] the percentage of mobile phone users in the world is around 78.05%. SMS is an ideal way of communication. Various studies manifest that an average person receives four times more SMS than e-mails.[2] Also, where an e-mail message has a probability of 20% getting opened, SMS comprises around 82% chance to be seen by the receiver at the end of the day. Around 18.7 billion text messages are sent over the network in one day.[3] Hence, the vast number of messages generated daily requires an automated mechanism to filter the undesirable and false messages from the necessary ones. Mainly, spam message comprises advertisements, offers, and winning lucky draws. Then, to proffer the rewards or prize amount, spammers request users to provide their details such as account number, one-time password, security pin, and other particulars, to steal the specifications and mug up the capital or finances of the person. Spammers bag a bounteous amount from malicious practices producing a significant loss to the authorized person or the organizations. Moreover, such unsolicited SMS occupies an ample capacity of message inbox. It may lead to skipping required or impromptu action messages that may incur a loss to the person. Further, these unwanted and toxic texts create unwanted traffic on both the server and client sides.

SMS spam is a substantial issue persuading in the world. People receive noticeable spam messages on their native SMS platform. Figure 1 shows the usage of different platforms preferred by the people and conveys that the use of native SMS platforms is to a great extent, besides the presence of other enduring applications. The figure illustrates the significance of chosen dataset graphically.

This research work aims at SMS classification as spam or ham, where spam refers to unsolicited messages and ham refers to legitimate ones. With the introduction of new technology and growth in the use of mobile devices, every day a large amount of SMS data is generated. The malicious activities add a significant aggregate of junk messages to datasets that must be eliminated. So classifying such a massive amount of data requires some automated procedures. Many ML models are available for classification, but the text messages mainly comprehend high-dimensional feature vectors. Incorporating such high-dimensional data to train the classifier is costly and with the presence of misleading and redundant data, the model may give erroneous results. Feature selection is an approach that can uncomplicate the categorization of spam messages and legitimate ones. It can overcome two constraints, i.e., model overfitting and the curse of dimensionality. Some features in the corpus are insignificant for training ML classification models and noisy data inculcated in the text misleads the classification model. Feature selection is a procedure to manually or automated select the features or the subset of features that can substitute the whole document.

This research contributes an ensemble feature selection algorithm that drops the irrelevant and redundant features from the dataset and assists the classifier in recognizing and releasing spam messages out of the user's inbox efficiently. We assemble feature engineering in the proposed ensemble algorithm employing hyperparameterized RFECV

[1] https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/
[2] https://financesonline.com/sms-marketing-statistics/
[3] https://shso.vermont.gov/sites/ghsp/files/documents/Worldwide%20Texting%20Statistics.pdf

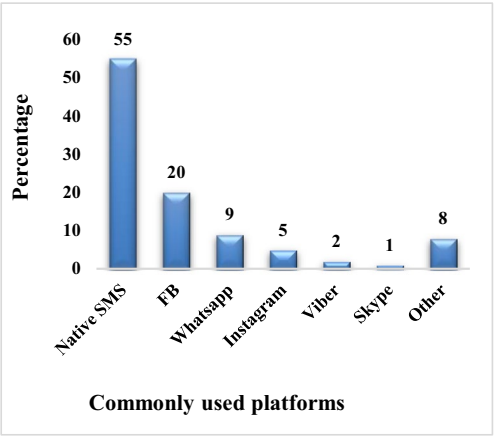


Fig. 1 Statistics of preferred messaging platforms (https://99firms.com/blog/sms-marketing-stats/#gref)

SVC and random forest algorithm. The rationale behind employing SVC and random forest algorithms is outlined in Section 6. The proposed algorithm reduces the computational cost by eliminating noisy and less effective features from the dataset. It mitigates the classifier in detecting spam and ham SMS producing the most optimal feature subset.

To illustrate the effectiveness of the proposed ensemble feature selection algorithm, the features subset is deployed to three ML models, including LR, SVM, and XGB. We compare the proposed ensemble feature selection algorithm with five other feature selection algorithms, including Chi-square, LV, PSO, MI, and RFE with all three classification models. Further, comparison of our proposed approach with the LSTM neural network model signifies the importance of the optimal feature subset obtained from the proposed algorithm. The study reveals that if the optimal feature subset is fed to the strong ML models, they give results analogous to the neural network models. This work also documents the motivation regarding stopwords removal from the dataset before classification by providing each classifier feature subset without removing stopwords and demonstrates results with all three ML classifiers. The experimental results are evaluated using f-0.5 score, precision, and recall.

The rest of the paper is organized as follows: Section 2 discusses the related work proposed by different authors for spam classification. Section 3 illustrates preprocessing techniques. Sections 4 and 5 present the literature review of utilized algorithms. Section 6 represents the proposed ensemble algorithm along with experimental work and results. Section 7 concludes the proposed work. Section 8 provides the future scope of work.

2 Related work

Recent years have witnessed a noteworthy growth in spam classification. This section describes different contributions by the researchers. Table 1 delineates the tabulated form of existing work. Abayomi et al. [1] reviewed seven databases, summarizing the research done in the field. The results illustrated that ML models like Naïve Bayes and SVM are preferred the most. Various statistics regarding the feature selection methodology, anti-spam application, corpora, performance metrics, and targeted approaches were calculated.

Table 1 Brief description of existing work

Ref	Proposed model	Description	FE	FS	Limitations	Research domain
Abayomi et al. [1]	Review paper	Deep analysis of the existing search strategies for spam SMS classification and emphasizes on exploring optimization algorithms	_	_	It considered only English publications	SMS spam
Sjarif et al. [30]	SVM classifier	Performance of SVM classifier was examined based on splitting proportions of dataset among train and test instances	StringToWordVector	Information gain	Performance metrics used are not justified for selected imbalanced dataset	SMS spam
Xia and Chen [34]	Discrete Hidden Markov model	Language independent model utilizes word order for SMS spam classification	BoW	_	Large training dataset vocabulary is required for classification	SMS spam
Bosaeed et al. [9]	Data filtration before classification	Proved effectiveness of SVM model employing five preprocessing and feature extraction filters. Model is dedicated to outgoing SMS	PF1-PF5	PF1-PF5	Optimization of software in fog, edge, and cloud layers may be done	SMS spam
Roy et al. [27]	Deep learning models	Illustrates the efficiency of hyperparametrized CNN and LSTM models	_	_	High computations for neural networks model	SMS spam

Table 1 (continued)

Ref	Proposed model	Description	FE	FS	Limitations	Research domain
Agarwal and Kumar [3]	Integrated Naïve Bayes and PSO model	Integrated Naïve Bayes algorithm, whose classification is optimized using PSO approach	BoW	Correlation-based method	Variants of Naïve Bayes and other optimization algorithms may be analyzed	E-mail spam
Kumaresan and Palanisamy [17]	Integrated SVM and S-cuckoo search model	Optimized S-cuckoo search algorithm for feature selection and deployed the dataset to SVM classifier	SCS	SCS	Other strong classification models can be considered for comparative evaluation	E-mail spam
Sjarif et al. [29]	Tf-Idf and random forest	Verifies the efficacy of Tf-Idf feature extraction algorithm and random forest classifier	Tf-Idf	_	More data preprocessing steps can be performed	SMS spam
Baaqeel and Zagrouba [7]	Hybrid model	Model enhances the detection rate by combining supervised and unsupervised algorithms	Tf-Idf	_	More preprocessing steps can be included and other classification models could be evaluated	SMS spam
Poonkodi and Sukumaran [23]	EASCEM model	Features selected through firefly optimization algorithm and fed to sequential adaboost classifier	WordNet	EFOA	Comparative evaluation with other feature selection algorithms could be done	E-mail spam
Makkar and Kumar [19]	PROTECTOR model	Utilizes deep learning algorithms centered over link properties and metadata of images	_	_	More hyper-tuned layers could be added to classifier and RNN models can be utilized for text features	Image spam

Table 1 (continued)

Ref	Proposed model	Description	FE	FS	Limitations	Research domain
El-Alfy and AlHasan [13]	DCA based framework	Signals for DCA framework are generated through two ML models (SVM and Naïve Bayes)	_	_	Signal generation with other robust ML algorithms can be analyzed	E-mail and SMS spam
Prilepok et al. [24]	Information fusion framework	Bayesian model is fused with signature tree and data compression algorithms for better performance	_	_	Amalgamation of feature engineering can be considered	E-mail spam
Ahmad et al. [4]	Feature engineering	Effectiveness of preprocessing steps and manual feature extraction. Categorizes the dataset among five classes comprising 28 sets	_	RFE	Evolutionary computing could be combined with ML models	Twitter spam
Adewole et al. [2]	SMSAD framework	Performs content-based Spam detection and spam account detection utilizing graph features of the network	_	Evolutionary algorithm	Ensembling content and network features can produce a more robust model	SMS and Twitter spam
Bhuvaneshwari et al. [8]	ACB model	Self-attention based CNN-BiLSTM deep learning model utilizes semantic contextual information for classification	_	_	Transfer learning and optimization algorithms could be inculcated for multi-level predictions	Spam reviews

Table 1 (continued)

Ref	Proposed model	Description	FE	FS	Limitations	Research domain
_	Proposed model	Ensemble feature selection algorithm produces 87% reduced feature set. Make ML models' efficiency analogous to neural network models	BoW	Ensemble model	_	SMS spam

Authors concern the limitation of dataset availability for spam classification due to privacy leakage and provide a publicly available dataset. Sjarif et al. [30] utilized SVM model to perform spam classification over the dataset retrieved from the UCI machine-learning repository. The authors evaluated results on the basis of different proportionality splits of the dataset among train and test samples. The StringToWordVector technique was used for feature extraction and information gain for feature selection. The results were explored with SVM, Naïve Bayes, and K-nearest neighbor algorithms along with their variants; SVM outperformed others. Xia and Chen [34] proposed a discrete Hidden Markov model based on the order of words for spam classification. The model was trained and tested over a dataset obtained from the UCI repository.

Table 1 particularize work done for spam detection in recent years. It includes Proposed model, Description; key notion of the study, Feature Extraction (FE) and Feature Selection (FS) techniques used in the research, Research domain specifying the area examined by the authors and Limitations specifying the enhancements that can be considered for future work in conjunction with the proposed approach.

The results were validated against a specified Chinese dataset that illustrated the model performed irrespective of the language. The model has limitations; it cannot work well for small datasets. Bosaeed et al. [9] studied spam detection approach that appealed to the fog, edge, and cloud layers (for server-side). The authors utilized the proposed model for outgoing messages. Three classification models were deployed, i.e., Naïve Bayes, SVM, and Naïve Bayes multinomial with five preprocessing and feature extraction filters (PF1–PF5). The work was evaluated over 15 datasets. The PF5 filter and SVM model yield the best outcome. Roy et al. [27] utilized neural network to classify the messages as spam or ham. The authors assessed hyperparameters tuned Convolution Neural Network (CNN) and LSTM. The results were compared with various ML models: Naïve Bayes, random forest, gradient boosting, LR, and stochastic gradient boosting. CNN model outperforms others with an accuracy of 99.4%. Agarwal and Kumar [3] proposed an integrated model Naïve Bayes with PSO for spam e-mail classification. Feature selection was performed using the correlation method over the Ling spam dataset. Naïve Bayes classifier was employed for initial classifications; thereafter, the results obtained were optimized with PSO algorithm. The integrated model outperforms the base model. Kumaresan and Palanisamy [17] utilized lemm-ling spam and bareling spam datasets for e-mail spam classification. A modified cuckoo search algorithm termed as step-size cuckoo search (SCS) has been implemented that performs feature selection to reduce the dimensionality of data and make classification more efficient. The fitness function calculation and classifications were done using the SVM model incorporating linear, polynomial, and quadratic kernels. The proposed model produces more accurate results than the SVM model with original cuckoo search algorithm. Sjarif et al. [29] used Tf-Idf method to extract the features from the dataset. The authors compared five ML classification models (Multinomial Naïve Bayes, KNN, SVM, decision tree, and random forest) to classify the text message as spam or ham. Among all the algorithms, random forest outperformed others with 97% f-measure, 97.5% accuracy, and 98% precision. Baaqeel and Zagrouba [7] proposed a hybrid system utilizing supervised and unsupervised ML algorithms. The authors preferred Tf-Idf representation for extracting feature space. The dataset obtained from the UCI repository was fed to selected six supervised ML models: Naïve Bayes, SVM, adaboost, LR, KNN, and random forest. The best results were obtained with SVM classifier. Then the unsupervised model (k-means clustering) was fed within the hybrid model along with a selected supervised ML algorithm (SVM) that outperformed other base models. The work avails

advantage of both supervised and unsupervised algorithms for spam message classification. Poonkodi and Sukumaran [23] proposed an effective ensemble model EASCEM for spam e-mail classification. It comprises of adaboost sequential classifier that performs differential weighting for the misclassified instances provided by the base model. The base learner gets the features selected through the Effective Firefly Optimization Algorithm (EFOA). Two datasets: enron and spam, were opted for this research work. The proposed algorithm provides compelling results in less time when compared with SCA and ALO-Boosting algorithms over both datasets. Makkar and Kumar [19] proposed a framework termed PROTECTOR; instead of focusing only on textual or image features, the model extends its concentration over link properties and metadata of images to classify them as spam or ham. The deep neural network has been utilized in this research and the optimal objective function is based on genetic algorithm. The experiments were performed over the image spam hunter dataset and achieve an accuracy of 96%. El-Alfy and AlHasan [13] proposed an approach for spam classification inculcating five datasets (two SMS datasets and three Spam e-mail datasets). The effectiveness of the datasets was compared and analyzed. Among them, the top two feature sets were opted. The Dendritic Cell Algorithm (DCA), which is based on immune system dendritic cells, is utilized in this study. DCA in amalgamation with two ML models (Naïve Bayes and SVM) had been exploited by authors for generating signals for classification. Prilepok et al. [24] proposed a method for spam e-mail detection fusing Bayesian filtering, signature trees, and normalized distance data compression algorithms. The model achieves good precision of about 99% for spam detection and 98% for ham detection outperforming the base models. However, the algorithms were found to be susceptive to training dataset framing. Ahmad et al. [4] presented a method for spam tweet detection. The study focuses on feature extraction and preprocessing techniques. The preprocessing step helps eliminate unused or less important words from the dataset. Features were distributed in five classes comprising 28 content-based feature sets. SVM classifier with two kernel variations (Gaussian and polynomial) employed for classification of tweets spam. The proposed model outperformed other models over the honeypot dataset (as stated in the work). Adewole et al. [2] proposed a unified framework for spam SMS and spam account detection. Four datasets were inculcated in this work. Categorization and reduction of features were performed to distinguish spam messages and accounts. Spammer's network was studied by exercising graph-based methods that helped in classification. Ten different ML algorithms were utilized to prove the efficacy of the proposed model, among all random forest outperforms others. Bhuvaneshwari et al. [8] dispensed an architecture for review classification based on CNN and Bi-LSTM model, i.e., self-attention CNN-based Bi-LSTM (ACB). It utilizes semantic contextual information for classification through document vector representation. The proposed framework outperforms its variants CNN and Bi-LSTM (CNNB), Attention mechanism and CNN (ACNN), CNN, Hierarchical Supervised Learning (HSL), and Spiral Cuckoo Search (Spiral CS). The authors also suggested necessitating multi-level predictions for spam detection.

The spam detection field gathered considerable attention from researchers in recent years, either for SMS, e-mails, reviews, profiles, or others. Existing models use deep learning algorithms, integrated models, and optimization algorithms. Feature reduction must be a requisite to handle such high-dimensional datasets as it can make the computations efficient and inexpensive. Although various authors have adopted feature engineering for spam detection, this study proposes two-level feature selection through a hyperparametrized

procedure capable of large-scale feature elimination producing the most optimal features subset that mitigates classification models.

Contributions of this paper are summarized as follows:

1. Proposed an ensemble feature selection algorithm that produces the most optimal feature subset. It eliminates approx. 87% of irrelevant and redundant features and enhance the classifier's predictions.
2. With optimal feature subsets, strong ML models can efficiently produce results analogous to deep learning models.
3. The purpose of stopwords removal is justified.
4. The proposed approach can serve as a baseline model that substitutes a high-dimensional dataset with a reduced optimal feature subset.

The performance metrics that can uphold the unbalanced dataset are applied to evaluate models, i.e., f-score, precision, and recall. According to recent work, SVM and ensemble models are strong models for spam detection. We adopt these models to establish the accomplishment of proposed feature selection method with most preferred classification models and the outcomes support the novelty of this research work. Further sections in this paper will outline our novel ensemble feature selection algorithm. Before producing a dataset for any classification model, some preprocessing is required which is discussed in the subsequent section.

3 Data preprocessing

This section discusses various data preprocessing techniques, from fetching the dataset to its conversion to the pattern deployed to feature selection or classification models.

Preprocessing techniques The selected corpus comprises text messages that must be preprocessed before feeding to the machine-learning models. Preprocessing techniques remove undesired features from the dataset. Several data preprocessing techniques deployed over the chosen dataset are described below:

(a). **Stopwords removal**

Stopwords are the usually utilized words present in both spam and ham messages, so it is assumed that they do not convey much useful information to our classification model. NLTK library provides an inbuilt library of stopwords in 'English'.

In this study, experiments are also conducted by training ML models with incorporated stopwords features in order to support the elimination of stopwords. The results of the experiments are provided in Section 6.

(b). **Stemming**

Stemming is the process of reducing inflectional structures or derivationally related morphological variants of words to their root words. We utilize "Porter Stemmer," which

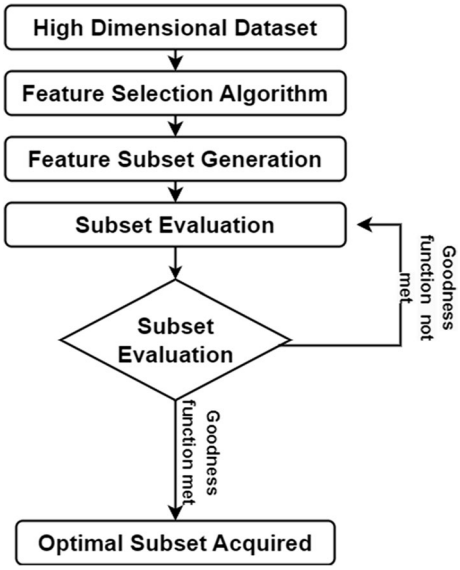


Fig. 2 Flow diagram for feature selection

works in five phases of elimination put in sequentially. It helps to determine vocabulary for the domain more expeditiously.

(c). **Bag of Words**

Feature extraction helps us to produce the numerical representation of the textual data. We represent our dataset using the Bag of Words (BoW) technique. The text or document is considered a bag (multiset) of words. It does not give importance to the order of the words in the corpus nor the grammar applied but keeps the multiplicity.

Keeping in view the contended implementation and further flexible customization of the fixed length vectors offered by the BoW algorithm, it is selected in this work.

After specific preprocessing steps, feature selection algorithms are applied to the dataset to obtain a more convenient feature subset. The feature selection algorithms utilized in this study are described in the next section.

4 Feature selection

The datasets acquired from numerous available sources over the network or assembled by the researchers for their experiments usually have a large number of features that makes it difficult for the machine-learning models to classify the data efficiently. Three types of features exist in the dataset: (i) Relevant features; that convey essential information to the classifier which is helpful in classification. (ii) Irrelevant features; that are useless for our classification model and may mislead the classifier declining their accuracy. (iii) Redundant features; that are needful to the classifier but the information contained in them had already been strongly conveyed through other features. The aim is to select features more relevant to the output target class and less correlated with other input features. Figure 2 shows the flow chart for feature selection.

The feature subset selection is a multi-objective goal, i.e., to balance the trade-off between two incompatible tasks: the first is to reduce the dimension of the dataset and the other is to increase the accuracy of the classifier [6]. The feature subset selection relies on some measures like distance, information present, dependency among input variables and with target output variable, and consistency [11].

The feature selection method is further categorized as wrapper, filter, and intrinsic (or embedded). Table 2 briefly describes all three categories of the feature selection algorithm.

This research proposes an ensemble feature selection algorithm and compares it with chosen feature selection algorithms; LV, Chi-square, MI, PSO, and RFE.

(a). **Low variance**

Low Variance (LV) is an unsupervised statistical feature selection technique. It considers only the input feature set, not the target output value. It eliminates features that have variance less than some predetermined threshold value. For instance, if a word is present in all the messages, it does not convey any differentiating potential to the spam or ham message classifier. Such features are said to have 0 variance. So, these insignificant features should be dropped. The given Eq. (1) provides the mathematical expression for variance [18].

$$variance\ score = p(1-p), \tag{1}$$

where, p represents the percentage of samples having the corresponding feature.

(b). **Mutual information**

Mutual Information (MI) of two arbitrary variables estimates their mutual relation. It is a technique evaluating the weight of information achieved for one random variable by studying another.

It is an application of information gain. This statistical technique not only considers the measure of interconnection between the features and output variable but also tries to find the relevance among the features in the dataset to reduce or eliminate redundant features. The scoring function for each feature is calculated as [18]

$$score\left(X_i\right) = I\left(X_i : Y\right) - \beta \sum_{X_j \epsilon S} I(X_k : X_j), \tag{2}$$

where, X_i is the feature for which we are calculating the mutual information score, $I\left(X_i : Y\right)$ is the relationship between the feature and target output variable, and $\beta \sum_{X_j \epsilon S} I(X_k : X_j)$ represents the correlation between the considered feature and selected feature subset.

(c). **Chi-square** (χ^2)

The chi-square test contributed by Karl Pearson in 1900, is a hypothetical statistics test. Here, it is utilized for feature selection.

The chi-square test for feature selection examines the dependency among stochastic variables in the provided dataset. It tries to evaluate the reliance of the target output

Table 2 Brief description of feature selection methodologies

	Definition	Advantages	Disadvantages
Filter-based	This method selects the subset of data by analyzing it through various statistical models to discover the relationship between the input features and the target output variable. It mainly uses some selection methods to obtain the input features with specified maximum values. Example: Chi-square, Euclidean distance, correlation-based	• Fast • Scalable • Independent of classifier	• Do not interact with the classifier
Wrapper- based	These methods calculate the importance of the feature subset by using the classifier. They produce the optimal feature subset utilizing a classifier. Example: Forward feature selection, Backward feature elimination	• Derives feature dependence • Interact with the classifier	• Results are delayed
Embedded	These are also termed hybrid models as they exercise the concept of both wrapper and filter-based feature selection techniques. They are implemented in models consisting of their own built-in feature selection procedures Example: Lasso, Ridge	• Accounts dependency among features • Interacts with classifier	• High cost

feature over the feature set. The feature subset is acquired by calculating the score for each feature given by Eq. (3) [18]:

$$\chi^2 = \frac{(observed\ frequency - expected\ frequency)^2}{expected\ frequency}. \tag{3}$$

The features having low χ^2 values are discarded as they reflect the uninformative behavior towards the classification or the target output value, and features with high χ^2 values are retained for the classification models.

(d). **Recursive feature elimination**

Recursive Feature Elimination (RFE) is a wrapper-based feature selection technique. It attempts to eliminate features with higher dependency and collinearity. It incorporates a classification model that removes the features with low ranking initiating with the whole dataset and repeats the procedure recursively until the determined aggregate is reached. Most of the time, this limit where to stop is unknown, so the default value of half of the features is set as stopping criteria. Numerous options are available to select the classification model, like SVM, perceptron, random forest, decision tree, and stochastic gradient boosting. RFE computes scores for different feature subsets and returns the more appropriate one.

It has a variant with an implicitly incorporated cross-validation approach, i.e., Recursive Feature Elimination with Cross Validation (RFECV). It selects the feature subset iteratively on ranking basis.

(e). **Particle swarm optimization**

Particle Swarm Optimization (PSO) is a metaheuristic algorithm proposed by Kennedy and Eberhart in 1995 [16] that uses an objective function to find the minimal optimal solution.

PSO is based on the food-searching nature of animals, like a flock of birds [33]. The birds seeking food are not aware of the location of food, but they can sense how far they are in each direction. The flock follows the bird with the least distance to the food.

In PSO, member of the population (flock) is referred to as particle (birds), each particle represents a solution, and their population is termed as swarm. Each member of the population is associated with velocity and position. The algorithm starts with some arbitrarily chosen population size that heads in some random direction throughout the search space. However, each particle remembers its own and the swarm's current best position. PSO updates the value in each generation to find the best optimal solution among all existing ones.

For a D- dimensional space, vector

$$X_i = X_{i1}, X_{i2}, X_{i3} \cdots X_{id}, \tag{4}$$

expresses the d-dimensional positional vector for the i^{th} particle and

$$V_i = V_{i1}, V_{i2}, V_{i3} \cdots V_{id}, \tag{5}$$

represents the d-dimensional velocity vector for the i^{th} particle.

For each iteration, the value of X_i and V_i are updated as follows:

$$V_{id}^{t+1} = w \times V_{id}^{t} + c_1 \times r_1 \times \left(p_{id} - X_{id}^{t}\right) + c_2 r_2 \times \left(p_{gd} - X_{id}^{t}\right), \tag{6}$$

$$X_{id}^{t+1} = X_{id}^{t} + V_{id}^{t+1}, \tag{7}$$

where, p_{id} represents the personal best for the particle and p_{gd} denotes the global best. c_1 *and* c_2 are learning factors. w indicates inertia weight. $r_1 and r_2$ are random variables lying under the range of [0,1]. Here, the velocity is also restricted to the predetermined valueV_{max}, i.e.,$V_{id}^{t+1} \in [-V_{max}, V_{max}]$.

We exert Pyswarm[4] Library [21] in this work, the fitness function of which is provided [32] as:

$$f(X) = \alpha(1 - P) + (1 - \alpha)\left(1 - \frac{N_f}{N_t}\right), \tag{8}$$

where, P is the classifier's performance, N_f is the feature subset, N_t represents total features, and α is the hyperparameter defining the trade-off between P and N_f.

Though the algorithm does not guarantees to come up with the best optimal ever. With the hinge of an idea, researchers are deploying the algorithm in various fields [28, 33]. It is utilized here for feature selection. Many variants of PSO are present, and we adopt discrete PSO.

After obtaining the feature subset from feature selection algorithms, the pruned dataset is deployed to classification models whose description is provided in the following section.

5 Classification techniques

The main aim of our research is to classify the given SMS as either spam or ham. Mainly classification techniques can be categorized in three ways: Machine-learning algorithms that learn from the given training dataset and construct a knowledge base using which it performs classification (LR and SVM), Bio-inspired algorithms based on the working of the human brain (LSTM), and Ensemble models that average the results obtained from multiple models (XGB). A brief outline of all the chosen models is provided below:

(a). Logistic regression

Logistic Regression (LR) [14, 20, 25] is a probabilistic statistical model. Depending on outcomes, they are of three types: Binomial, Multinomial, and ordinal. Here, we utilized binary LR. Binary LR is based on linear separation available among the data points. It aims at locating the most optimal model, delineating the association among high-dimensional independent feature vectors and targeted dichotomous dependent variable.

It uses the logit or sigmoid function, whose mathematical expression is given by Eq. (9). It maps the real values in the range [0, 1].

$$g(z) = \frac{1}{1 + e^{-z}}, \begin{cases} z \to \infty, g(z) \to 1 \\ z \to -\infty, g(z) \to 0 \\ z = 0, g(z) = 0.5 \end{cases} \tag{9}$$

[4] https://pyswarms.readthedocs.io/en/development/examples/feature_subset_selection.html

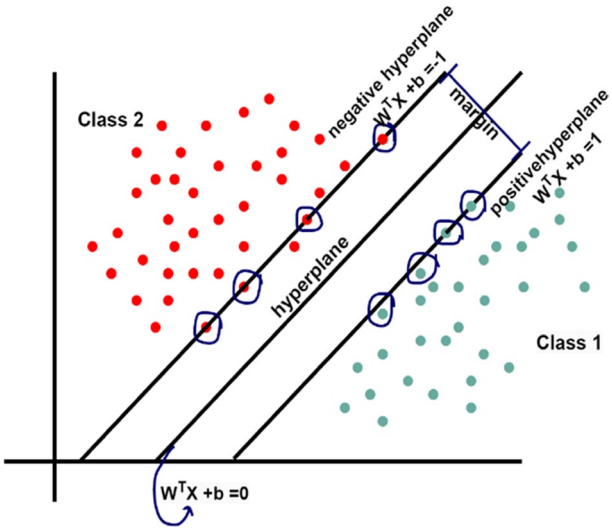


Fig. 3 Visual description of SVM classifier

The sigmoid function's derivative is elementary, making it attractive to use. Its derivative is computed as:

$$g\prime(z) = g(z)(1 - g(z)). \tag{10}$$

Let us have the target variable t, whose output value can be either 0 or 1, and X be the input variable of independent features.

Then the probability for *'t'* provided X is calculated as

$$P(t|X) = \begin{cases} (1 - h(X))^{1-t}; & for\ t = 0 \\ h(X)^t, & for\ t = 1 \end{cases} \tag{11}$$

where, $h(X) = g\left(\beta^T X\right)$ and β is the regression coefficient vector whose optimal values are learned using the maximum likelihood approach. The likelihood of β is expressed as

$$L(\beta) = \prod_X P(t/X;\beta). \tag{12}$$

The logarithmic odds (log function) of above Eq. (12), $l(\beta)$ (which is equal to $\log(L(\beta))$) is

$$l(\beta) = \sum_X t\ \log\left(h(X)\right) + (1 - t)\log(1 - h(X)), \tag{13}$$

and the cost function of the algorithm, which is inversely proportional to the likelihood parameter is given as:

$$cost(\beta) = \sum_X -t\ \log\left(h(X)\right) - (1 - t)\log(1 - h(X)). \tag{14}$$

The algorithm aims at maximizing the likelihood and minimizing the cost.

(b). **Support vector machines**

Support Vector Machine (SVM) is the most applied technique to stratify mail as junk or essential [31]. It is a supervised ML algorithm that can be utilized for regression or

classification. Training the SVM model is easy and outperforms several other ML models [26, 31]. It uses information from the high-dimensional database. Figure 3 provides the graphical representation of the SVM classifier with two-class instances. SVM is a statistical model that limits the risk in classification by defining support vectors. It tries to isolate the classes by constructing a hyperplane through these vectors. The support vectors are highlighted with a circle over the negative and positive hyperplanes in Fig. 3. The discriminant function for them is given below:

$$g(X) = W^T X + b = \sum_{i \in SV} \alpha_i X_i^T X + b, \tag{15}$$

where, SV denotes the support vectors. X is the test sample. The function relies upon the dot product between X and the support vector X_i. W represents the linear combination of the data points. Classifier tries to expand the margin between support vectors and obtained hyperplane. The algorithm does not rely on the likelihood distribution of data in the corpus, and it straightforwardly fits the classes in the dataset.

As it is not always possible that the data obtained will be linearly separable for such situations, the notion of “kernel function" is utilized. For non-linearly distinguishable data points, some feature transformation is performed that transforms the data points to higher dimensions so that the decision boundaries for the extrapolated feature space can be found. Such manual conversion increases computation costs as it usually deals with high-dimensional feature space. However, using kernel functions it is achieved without significant implications on the computation cost. SVM reduces the issue of overfitting and can handle high-dimensional data [12], but choosing the appropriate kernel function is monotonous work.

(c). Extreme gradient boosting

Extreme Gradient Boosting (XGB) [10], initiated by Deep Machine Learning Community (DMLC) as a research project, gained popularization after its success in the Higgs machine learning competition. It is an ensemble learning technique [15] that enhances the performance of ML models. Two techniques to ensemble or combine the models: Bagging and Boosting. Bagging (Bootstrap aggregating) tries to enhance predictions by making a better-combined classifier through amalgamation of the predictions given by a single model. Boosting (Adaptive resampling and combining) ameliorates the accuracy of the predictions by frequently running the base model over different training data distributions. Ensemble models are obtained by combining weak or base models to accomplish higher precision than single models.

XGB is a boosting ensemble model and a competitive spam classifier [10, 22]. It is a modification of the gradient-boosting algorithm that constructs the decision trees sequentially. Weights are given high significance in this algorithm. All the independent variables are associated with some weights, which are subsequently put into decision trees that predict the output. The weight of the variables mispredicted is raised and those variables are then assigned to another decision tree. These single predictors are combined to create a more robust and precise model. XGB is highly scalable, can utilize parallel processing, and is faster than the gradient boosting algorithm.

(d). Long Short Term Memory

Bio-inspired networks have been getting more attention since the last decade because of their accuracy. Long Short Term Memory (LSTM) [15] is the Recurrent Neural Network

(RNN), where the computation includes the present input state and the results from the previous state by accumulating hidden layers. It can remember long sequences on account of which it is used for many applications like language models, speech, and video data streams. It consists of four gates (forget gate, input gate, memory gate, and output gate):

Forget gate(f_t) helps the LSTM model decide what information needs to be retained from the previous state. Its value lies between [0,1] and 0 signifies the discardment of the complete information.

$$f_t = \sigma\left(W_f.\left[h_{t-1}, x_t\right] + b_f\right), \tag{16}$$

where, b_f represents the bias, h_{t-1} is the previous state output, W_f is the weight, and x_t is the current input. 't' represents the current state while '$t+1$' and 't-1' indicates the next and previous states.

Then, to decide what information about the current state should be stored is done through two gates. The updation of values is performed through the input gate (i_t) and the tanh layer gives a vector $\left(\widehat{C}_t\right)$ that is added to the input layer.

$$\widehat{C}_t = \tanh(W_C.\left[h_{t-1}, x_t\right] + b_C). \tag{17}$$

$$i_t = \sigma(W_i.\left[h_{t-1}, x_t\right) + b_i. \tag{18}$$

Further, it updates the memory gate from the previous state C_{t-1} to the current state C_t as

$$C_t = f_t \times C_{t-1} + i_t \times \widehat{C}_t. \tag{19}$$

Then, it scales the output (o_t), and the filtered output (h_t) given by the tanh layer

$$o_t = \sigma\left(W_o.\left[h_{t-1}, x_t\right] + b_o\right). \tag{20}$$

$$h_t = o_t \times \tanh\left(\widehat{C}_t\right). \tag{21}$$

LSTM models have certain advantages, like relative insensitivity of data intervals and solving vanishing gradient problem. Still, they have certain limitations: more training time, large memory, and difficulty implementing dropout.

6 Proposed model and experiment work

This section describes the proposed feature selection ensemble algorithm and gives a determined overview of experiments proving the efficacy of the proposed approach.

6.1 Dataset

The dataset is collected from UC Irvine (UCI) machine learning repository. It was established as an ftp archive in 1987 by David Aha and fellow graduates[5] to help the researchers' community.

[5] https://archive.ics.uci.edu/ml/about.html

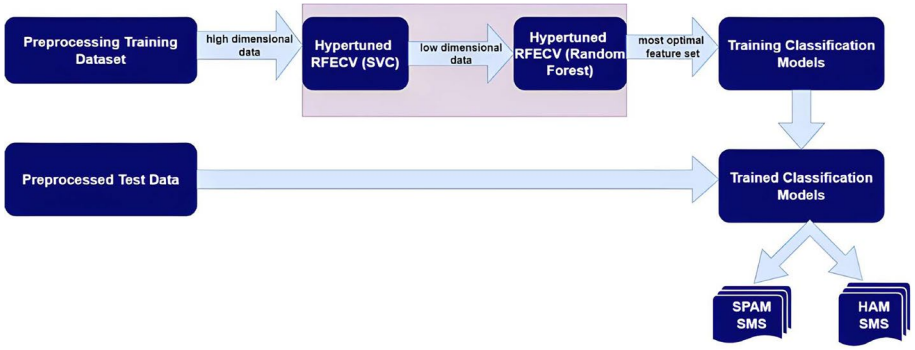


Fig. 4 Block diagram for the proposed ensemble feature selection algorithm

This research is availed on the SMS spam database [5]. The dataset consists of 5574 SMS samples, with 13.40% (around 747) spam and 86.60% (around 4827) legitimate messages. Here, the spam messages are labeled as 'spam' and those unsolicited as 'ham'. As per token statistics, 63,632 are ham, while 17,543 are spam. Examples of spam words in this dataset: free, winner, claim, reward, cash, jackpot, congrats, urgent, bonus, account, private, unredeemed, sex, contact, and vouchers.

6.2 Proposed algorithm

It is estimated from the statistics selected dataset is an imbalanced dataset that resembles the realistic distribution of data present in the environment. As there are a large number of features present in the dataset to reduce such a high-dimensional dataset, this paper proposes an ensemble feature selection algorithm that helps classification models produce exceptional results with pruned and precise feature subset. The presence of irrelevant features mislead the classifier and the redundant features lead to expensive computations. To eliminate such features, the proposed approach exploits RFECV wrapper-based feature selection algorithm using two hyperparameters tuned models: Support Vector Classifier (SVC) and random forest classifier. The novel ensemble feature selection algorithm gives fitter solutions with just 13% of the selected optimal feature subset. Figure 4 presents the block diagram of the proposed algorithm. The diagram provides a brief visual perspective of the ensemble feature selection algorithm proposed in this work.

6.3 General Remarks

(i) *Significance of applying SVC classifier before random forest algorithm*

RFECV is a wrapper-based feature selection algorithm that provides a subset of features substituting the high-dimensional dataset. Initially, we deploy the preprocessed dataset to hyperparameters tuned RFECV with SVC classifier as SVC is a scalable classifier and memory-efficient algorithm that can handle high-dimensional data efficiently and

cost-effectively. Then selected subset is deployed to hyperparameters tuned RFECV with random forest classifier as random forest algorithm is the bagging method that selects the instances of features set provided to it and not all at once or attribute bagging. Deploying a large feature set to the small number of decision trees in the random forest algorithm will result in the omission of some features or they will be tested only once and conveniently increasing the number of trees will amplify computational cost. So, to avoid such complications cost-effectively and efficiently, hyper parameterized RFECV with the random forest is fed with the subset of selected features produced by RFECV with hyperparameters tuned SVC.

(b) *Purpose of Stratified K-fold Cross-Validation(SKCV)*

Instead of using K-fold cross-validation, we employ SKCV, as it untangles the problem of random sampling and overcomes class-imbalance dataset generalization errors. The algorithm is validated over each sample while selecting features subset using SKCV.

(iii) *Consequences of comparing with the LSTM algorithm*

LSTM is selected based on its efficiency in learning long-term dependencies in data. It is a powerful RNN extension, combating the vanishing gradient problem. We compare classification results with LSTM to justify the obtained optimal feature subset's effectiveness analogous to a strong neural network with inexpensive computation.

(iv) *Grounds of applying feature selection over training dataset only.*

In this work, feature engineering is implemented only over the training dataset so that test samples remain unrevealed to the classifiers. It is done to make the model practical for real-world applications where the test samples are anonymous to the classifier.

The empirical study verifying the plump for hyperparameters required for the selected dataset, which can be extended to a similar high-dimensional realistic dataset prevailing over the network is as given below:

6.4 Hyperparameters tuned RFECV with SVC

SVC is a linear classifier. It classifies the sample among the given target functions or the class labels. Here, it is a two-class problem, i.e., either the sample will be "spam" represented through categorical value '1' or "ham" with categorical value '0'. Different hyperparameters are tested over the training dataset to set the value for the proposed algorithm. The interpretation of adopted standards for hyper-tuned RFECV with SVC is as follows:

(a). *Kernel:*

SVC is an exceptional algorithm in classification. It is a linear classifier having a mathematical function termed kernel. There are many kernel functions, including linear, radial basis function, polynomial, and others. Generally, the function that qualifies "Mercer's condition" can be used as a kernel function. Equation (22) gives Mercer's condition for any kernel $k(x, y)$ represented in higher dimensional space using the dot product.

$$\sum_{i,j} k\left(x_i, x_j\right) c_i c_j \geq 0. \tag{22}$$

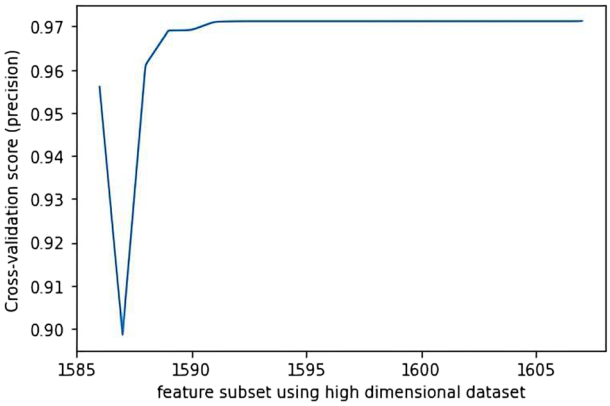


Fig. 5 Grading score on account of feature subset with hyper-tuned RFECV + SVC

We adopt a linear kernel given by Eq. (23). The appropriate kernel function is chosen on the dataset's foundation.

$$K\left(x_i, x_j\right) = x_i.x_j. \tag{23}$$

The kernel function measures the similarity between features x_i, x_j, mostly symmetric similarity. The choice for the "linear kernel" function is based on different measurement scores provided by the mathematical formulations, proving the linear classification stability of the selected dataset. Figure 5 depicts the graph line between the number of features selected with the SKCV 'precision score'. A similar graph is visualized through other performance metrics like: 'accuracy score' and 'recall'.

As performance metrics score depicted in Fig. 5 are contented and also the number of selected features remains related to accuracy metrics. So, it verifies that the data points are linearly separable. Consequently, the choice of "linear kernel" is verified.

(b). *Penalty parameter:*

SVC tries to maximize the margins and fit training data. The penalty parameter 'C' deals with noise in the margins of the classifier. It is used to control training errors. Some data points do not lie in the correct portion of the decision boundary. Such data points are entitled as slack variables (ξ). Equation (24) expresses the slack variables association with the parameter 'C', i.e., the penalty for each wrongly placed training sample.

$$w.w + C\sum_{k=1}^{m} \xi_k. \tag{24}$$

Experiments are performed with small to large random values of C ranging from 1 to 1000, and a slip in accuracy is found with larger values of 'C'. Thus with a decline in the score, it concludes that the data points need highly marginalized decision boundaries providing some degree of ignorance for the outliers. So, we keep the value for C equal to '1' or adopt to ignore some outliers as the random forest classifier will further sort them.

Table 3 shows all the parameters adopted for hyperparameters tuned RFECV with SVC.

6.5 Hyperparameters tuned RFECV with random forest algorithm:

Random forest algorithm, an estimator based on bagging ensemble learning. It contains defined decision trees applied to various subsets of the dataset whose predictions are averaged to classify the test sample among given class labels. The interpretations of adopted hyperparameters for the algorithm are as follows:

(a). *n_estimators:*

Random forest is an ensemble model using decision trees for classification. 'n_estimators' defines the number of decision trees whose predictions are amalgamated to obtain more reliable results. Notwithstanding the verity of stretch in accuracy with an increase in the number of trees, it also led to the construction of a slow and expensive computational model. Experiments are performed for 10, 30, 50, 80, 100, 150, 200 n_estimators and some more random numbers in between the range of 10–200. From the results obtained, it is concluded that 50 n_estimators produce a grading score approaching to stacked accuracy, as illustrated in Fig. 6.

The further addition of decision trees does not consummate the cost incurred with achieved gains. Figure 6 gives the plot for the SKCV grading score achieved using RFECV with random forest algorithm applied over the subset of features obtained after hyperparameters tuned RFECV with SVC feature selection algorithm.

(b). *Criterion:*

In decision trees, there are two criterion: Entropy and Gini impurity. Both criterion are helpful as they provide knowledge regarding whether the split is pure or impure in reference to the opted feature. An optimal split leads to a more accurate and less depth of the tree. Equation (25) represents Gini impurity and Eq. (26) entropy. Gini impurity is based on the selected feature's probable index measuring randomness of the misclassified sample and entropy measures information indicating the disorder of the opted features accompanied by the target. Entropy is a logarithmic mathematical calculation with costly computation whose output ranges from 0 to 1. At the same time, Gini impurity involves algebraic mathematical formulations, making it computationally inexpensive and its output ranges from 0 to 0.5. In both criterion, the value of '0' indicates most pure split and higher values signify the split's impurity.

$$Gini\ Index = 1 - \sum_j p_j^2. \tag{25}$$

$$Entropy = \sum_i -p_i log_2 p_i. \tag{26}$$

Table 3 Adopted hyperparameters For RFECV + SVC

Parameters for SVC	Parameters for RFECV
Kernel = linear, C = 1, probability = False	Estimator = hyperparameterized SVC, step = 0.05, cv = stratified K-fold(2), scoring = 'accuracy'

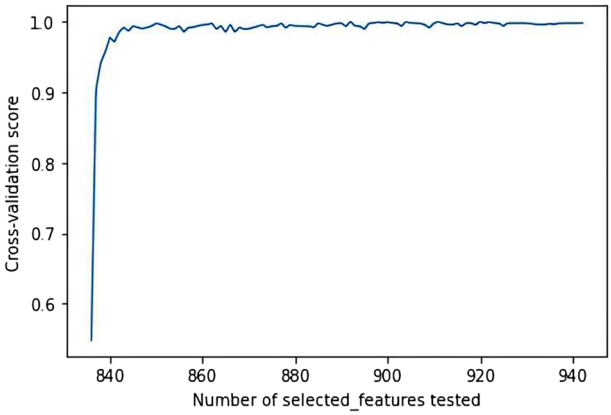


Fig. 6 Grading score on account of optimal feature subset with hypertuned RFECV + random forest

With the reduced dataset, we test for both the criterion and find that entropy is slightly accurate and provides more satisfying results than Gini impurity. Realizing the dominance of prediction results, 'entropy' is adopted over 'Gini impurity'.

Table 4 shows all the parameters adopted for hyperparameters tuned RFECV with random forest classifier.

The effectiveness of the ensemble algorithm is conveyed through Figs. 5 and 6, as the grading score for RFECV + SVC with approx. 25% features is 0.97 (or 97% accuracy achieved over training samples). After deploying those selected features to the RFECV + random forest algorithm, the cross-validation score is approaching 1(or 100% accuracy achieved over training samples), i.e., with almost half of the features (13%) as illustrated in Fig. 6. This reflects the compelling nature of the proposed ensemble algorithm that produces more acceptable and accurate results with few features. The results estimate the presence of irrelevant features that cannot be eliminated by isolated RFECV + SVC feature selection algorithm and justify the filtration of features at another level. Further nobility of the proposed ensemble algorithm is proved in the subsequent portion of this manuscript by comparing its result with the other five feature selection algorithms.

6.6 Experimental work and comparative study

Numerous experiments are performed to verify the efficacy of the proposed ensemble feature selection algorithm that supports the novelty of this research work. Figure 7 gives a brief visual outline of the comparative analysis.

Table 4 Adopted hyperparameters for RFECV + random forest classifier algorithm

Parameters for random forest algorithm	Parameters for RFECV
n_estimators = 50, criterion = 'entropy', min_sample_split = 2, min_sample_leaf = 1, bootstrap = True	Estimator = hyperparameterized random forest classifier, step = 0.01, cv = stratified K-fold(5), scoring = 'accuracy'

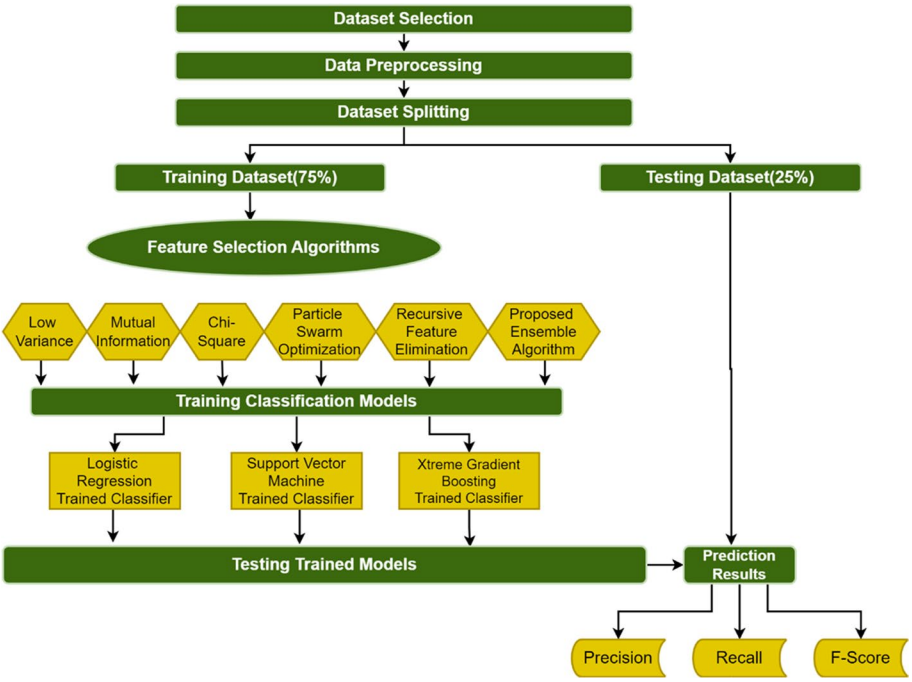


Fig. 7 Experiments outline

The proposed ensemble feature selection algorithm is differentiated by comparing with selected five feature selection algorithms: statistical Chi-square, LV, MI, RFE SVM, RFE with random forest classifier, and meta-heuristic PSO. The feature subset resulting from each algorithm is experimented with three classification models: LR, SVM, and XGB. Also, the LSTM neural network classifier with all features is computed and compared. Additionally, we feed each classification model with a dataset comprising stopwords to validate if the inculcation of stopwords can give some knowledge to the classifier that assists in classification. For all the experiments split of the training and test dataset is identical (the training dataset comprising 75% instances of the dataset and testing dataset containing 25% instances of the provided dataset). Feature selection is implemented over the training dataset only.

The performance metrics exploited to evaluate the classification models are recall, precision, and f-beta score, whose mathematical expressions are as given below:

$$Precision = \frac{TP}{TP + FP}, \tag{27}$$

$$Recall = \frac{TP}{TP + FN}, \tag{28}$$

Table 5 Performance metrics parameters

Actual values	Predicted values	
	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FP)	True Negative (TN)

Table 6 Number of features selected through different feature selection algorithms

Feature selection approach	Number of features selected
Total features with stopwords	6396
Total features without stopwords	6296
Low variance	5464
Mutual information	3228
Chi-square	3148
PSO	3778
RFE + SVC	3148
RFE + random forest	3148
Proposed ensemble algorithm	836

$$f-\beta\ score = \left(1+\beta^2\right)\frac{Precision * Recall}{\left(\beta^2 \times Precision\right) + Recall}, \tag{29}$$

where, β implies the value chosen for f-score. TP, FP, TN, and FN represent True Positive, False Positive, True Negative, and False Negative respectively. Table 5 illustrates the provided parameters.

In this work, FP is more threatening because the ham message mispredicted and put in the spam folder may hinder some prime information from the user. So, the value of β is 0.5 for the $f-\beta\ score$ concerning the importance of precision over recall.

6.7 Experimental results

Table 6 shows the number of features selected by the proposed ensemble feature selection algorithm, five selected feature selection algorithms and the total number of features in the dataset with and without stopwords.

The proposed ensemble feature selection algorithm produces a feature subset with the least features whose validity is justified through experimental work done and communicated in this research.

Optimal choices made for different selected feature selection algorithms are described as follows:

The attempt for feature selection using different selected feature selection algorithms is made with ideal choices.

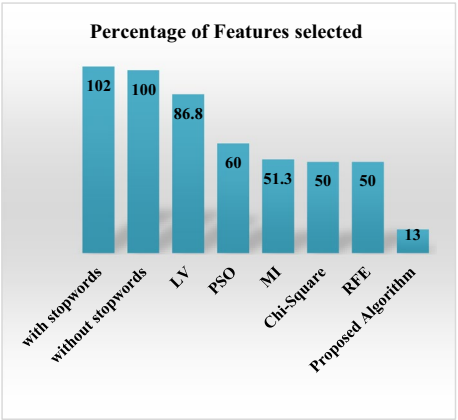


Fig. 8 Percentage of features selected using different feature selection algorithms

- For Chi-square, the choice of 50% is put as further decreasing the count of features the performance of the classification model degrades.
- With LV, all the features with zero threshold are eliminated.
- Using MI, all features with zero relevance are withdrawn.
- RFE with SVM and random forest disposes of the specific selected features subset.
- PSO is iterated 100 times with experimented and selected most optimal hyperparameters of the algorithm.

Figure 8 gives the percentage of features produced through different employed and proposed feature selection algorithms. Here, the feature set without stopwords is considered as complete or 100% features vector space(i.e., 6296 variables) while feature space with stopwords is presented as 102%, i.e., with extraneous features (or without stopwords elimination 6396 variables) and the rest of them are feature subset provided by respective selected feature selection algorithms.

The performance comparison of all three ML classification models on the feature set with stopwords, without stopwords, and feature subset selected employing five feature selection algorithms along with the proposed ensemble feature selection algorithm in terms of three performance evaluation metrics: precision, recall, and f-0.5 score is provided in Table 7. The potential of obtained optimal feature set obtained using the proposed ensemble feature selection algorithm on all three ML models is compared with the LSTM model also.

The acronym for Table 7 in the Models column is as follows The classification model name is suffixed by the features selection algorithm following the hyphen. The "SW" followed by the classification model name in the table indicates the model trained and tested embracing the stopwords features set. The bold data in table presents the results obtained when classifiers are fed with features selected through proposed ensemble feature selection algorithm. Figure 9 presents the bar graph for the percentage of features selected and f-0.5 score for all three ML models.

Spam filtering applications need to be more biased towards the precision score as the spam message misclassified as ham and put into the inbox is tolerable. However, a ham

Table 7 Results of experiments performed

Models	Precision	Recall	F-0.5 Score
SVM	94.61	87.78	93.16
SVM-SW	92.94	87.78	91.86
SVM-LV	95.53	92.43	94.89
SVM-MI	95.05	93.51	94.74
SVM-PSO	94.85	89.72	93.78
SVM-Chi-square	95.12	86.67	93.30
SVM-RFE (SVC)	95.53	92.43	94.89
SVM-RFE (random forest)	96.04	91.89	95.18
SVM-Proposed algorithm	**98.27**	**92.43**	**97.04**
LR	94.47	85.56	92.54
LR- SW	93.97	86.67	92.41
LR-LV	98.18	87.56	95.85
LR-MI	97.54	85.94	94.98
LR-PSO	98.13	85.40	95.29
LR- Chi-square	94.51	86.11	92.70
LR-RFE (SVC)	98.18	87.56	95.85
LR-RFE (random forest)	98.18	87.56	95.85
LR-Proposed algorithm	**98.19**	**88.11**	**95.99**
XGB	96.20	84.44	93.59
XGB-SW	97.50	86.67	95.12
XGB-LV	98.23	90.27	96.53
XGB-MI	96.93	85.40	94.38
XGB-PSO	96.31	84.86	93.78
XGB- Chi-square	95.67	86.11	93.59
XGB-RFE (SVC)	98.23	90.27	96.53
XGB-RFE (random forest)	98.23	90.27	96.53
XGB-Proposed algorithm	**99.41**	**90.81**	**97.56**
LSTM	98.79	90.27	96.88

message with emergency action requisite, misclassified and put in the spam folder will be precarious and unbearable.

The highlighted results in Table 7 specify the scores obtained when the proffered features subset from the proposed feature selection algorithm is fed to the ML models (LR, SVM, and XGB). It can be visualized from the results that the proposed algorithm with the least optimal features vector space when fed to all three ML models, provides better results than other feature selection algorithms. This concludes that the classification models give more acceptable results with reduced computations. The proposed novel ensemble feature selection algorithm can produce the most relevant and pruned optimal features subset among the different feature selection techniques. It helps the classifiers predict the class's label more accurately and with just approx. $1/10^{th}$ of computation.

It is depicted from the results obtained that the performance metrics with other feature selection algorithms are also quite acceptable, which signifies the absence of a large number of irrelevant features in the dataset. However, with just 13% of features more accurate results are acquired employing the proposed ensemble feature selection algorithm indicating the presence of approx. 87% redundant and some irrelevant features. So, the proposed

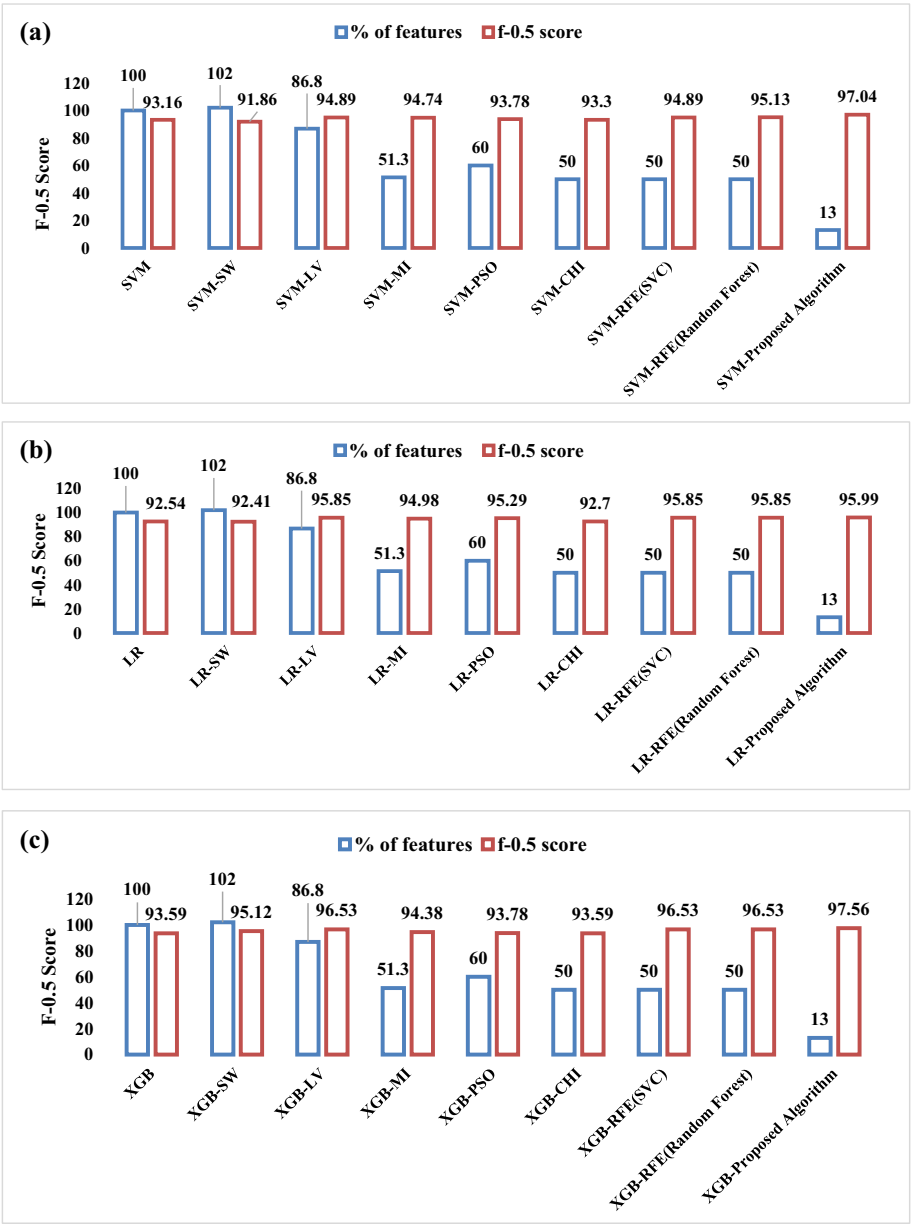


Fig. 9 **a** Bar graph for f-0.5 score and percentage of features selected for SVM model with all feature selection algorithms. **b** Bar graph for f-0.5 score and percentage of features selected for LR model with all feature selection algorithms. **c** Bar graph for f-0.5 score and percentage of features selected for XGB model with all feature selection algorithms

ensemble algorithm can be utilized for removing redundant features with justifiable two-level verified feature elimination procedures (RFECV + SVC and RFECV + random forest). Mostly the real datasets are also indulged with such high dimensions of redundant features

that need to be pruned as they occupy large memory and cellular phones with limited memory space found it baffling. This suggests feature selection approach is a requisite for spam classification and the percentage of features eliminated along with enhanced results rationalize the adoption of proposed ensemble approach for SMS spam classification.

The results conclude that all three models provide competitive results, but XGB gives slightly more satisfactory results. It can be accounted on the basis of advantages XGB distributed boosting algorithm provides. It is highly scalable and resists overfitting. It is a tree-based approach supporting regularization and apprehend dependencies more than the linear models. It can handle missing data, which mostly prevails in the SMS dataset. With all these adaptations towards the data, it is likely to produce a finer classification.

As depicted from the given results, the proposed algorithm with the least dimensionality features subset has given the best results and ML models outperformed LSTM. It is deduced that the optimal features subset obtained from the proposed algorithm, when fed to SVM and XGB classification models produces results comparable to the LSTM model. This illustrates that the optimal feature subset with reduced feature space when imparted to train strong ML models, produces results competitive to the bio-inspired or neural network models.

The study concludes the grounds of stopwords removal from the dataset, as they not only make computation expensive by increasing the dimensionality but also reduce the model's accuracy for classification. It is outlined in the given study that feature selection algorithms incorporated before classification improves the efficiency of the classifier with reduced feature space (inexpensive computation) and predictions are also enhanced. In closing, the proposed ensemble feature selection algorithm can be utilized for real-time performance.

7 Conclusion

With the drift of people and society towards the online world, a large proportion of the population possess mobile phones. SMS is one of the most preferred ways to communicate. However, along with all the advancements, some malicious activities have also arisen, and spam message is one such practice**.** The large volume of data emerges the need for automated models to filter such rogue data from the network. High dimensions of the dataset result in a computationally expensive classifier. The presence of trivial and irrelevant features deceives the classifier resulting in misclassifications. This study proposes a novel ensemble feature selection algorithm that prunes 87% of irrelevant or redundant features and performs more efficiently in comparison with other feature selection algorithms (LV, MI, PSO, RFE, Chi-square) over all the three classifiers (LR, SVM, XGB). The proposed ensemble feature selection algorithm with strong ML models surpasses the LSTM neural network model.

This research work illustrates the efficacy of an ensemble feature selection algorithm for high-dimensional datasets in text classification. The study also demonstrates that the stopwords convey no vital information to the SMS spam detection classifier. The proposed algorithm can serve as a baseline feature selection algorithm for spam classification or any related high-dimensional dataset. Experimental outcomes conclude that the proposed approach is preferable to avoid redundant features effectively from the dataset.

8 Future work

This research illustrates the importance of ensemble learning that can be enhanced with more hyperparameters tuning or inculcating different hybrid algorithms. This work can be further extended by incorporating images with text datasets for better results. Integration of ensemble learning with neural networks compelling cost-effective models could be interesting. Superior models call for a language-independent algorithm as the proposed approach is limited to English. Models detecting such unsolicited data on the server side would be beneficial. Much work is done for e-mail spam classification; considerable research is requisite for SMS spam classification.

Declarations

Conflict of interest We declare that we do not have any conflict of interest in the publication of this work.

References

1. Abayomi-Alli O, Misra S, Abayomi-Alli A, Odusami M (2019) A review of soft techniques for SMS spam classification: methods, approaches and applications. Eng Appl Artif Intell 86:197–212
2. Adewole KS, Anuar NB, Kamsin A, Sangaiah AK (2019) SMSAD: a framework for spam message and spam account detection. Multimed Tools Appl 78(4):3925–3960
3. Agarwal K, Kumar T (2018) E-mail spam detection using integrated approach of Naïve Bayes and particle swarm optimization. In: 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, pp 685–690
4. Ahmad SBS, Rafie M, Ghorabie SM (2021) Spam detection on Twitter using a support vector machine and users' features by identifying their interactions. Multimed Tools Appl 80(8):11583–11605
5. Almeida TA, Hidalgo JMG, Yamakami A (2011) Contributions to the study of SMS spam filtering: new collection and results. In: Proceedings of the 11th ACM symposium on document engineering, pp 259–262
6. Al-Tashi Q, Abdulkadir SJ, Rais HM, Mirjalili S, Alhussian H (2020) Approaches to multi-objective feature selection: a systematic literature review. IEEE Access 8:125076–125096
7. Baaqeel H, Zagrouba R (2020) Hybrid SMS spam filtering system using machine learning techniques. In: 2020 21st International Arab Conference on Information Technology (ACIT). IEEE, pp 1–8
8. Bhuvaneshwari P, Rao AN, Robinson YH (2021) Spam review detection using self attention based CNN and bi-directional LSTM. Multimed Tools Appl 80(12):18107–18124
9. Bosaeed S, Katib I, Mehmood R (2020) A fog-augmented machine learning based SMS spam detection and classification system. In: 2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC). IEEE, pp 325–330
10. Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp 785–794
11. Dash M, Liu H (1997) Feature selection for classification. Intell Data Anal 1(1–4):131–156
12. Dogan Ü, Glasmachers T, Igel C (2016) A unified view on multi-class support vector classification. J Mach Learn Res 17(45):1–32
13. El-Alfy ESM, AlHasan AA (2016) Spam filtering framework for multimodal mobile communication based on dendritic cell algorithm. Futur Gener Comput Syst 64:98–107
14. GuangJun L, Nazir S, Khan HU, Haq AU (2020) Spam detection approach for secure mobile message communication using machine learning algorithms. Security and Communication Networks 2020:1–6
15. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735
16. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks, vol 4. IEEE, pp 1942–1948
17. Kumaresan T, Palanisamy C (2017) E-mail spam classification using S-cuckoo search and support vector machine. Int J Bio-Inspired Comput 9(3):142–156

18. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, Liu H (2017) Feature selection: a data perspective. ACM Computing Surveys (CSUR) 50(6):1–45
19. Makkar A, Kumar N (2021) PROTECTOR: an optimized deep learning-based framework for image spam detection and prevention. Futur Gener Comput Syst 125:41–58
20. Menard S (2002) Applied logistic regression analysis, vol 106. Sage, New York
21. Miranda LJ (2018) PySwarms: a research toolkit for particle swarm optimization in python. J Open Source Softw 3(21):433
22. Mustapha IB, Hasan S, Olatunji SO, Shamsuddin SM, Kazeem A (2020) Effective E-mail spam detection system using extreme gradient boosting. arXiv preprint arXiv:2012.14430
23. Poonkodi, T., & Sukumaran, S. (2021). Effective Adaboost Sequential Classification Algorithm based Ensemble Method for E-Mail Spam Filtering. Annals of the Romanian Society for Cell Biology, 13056–13066
24. Prilepok M, Berek P, Platos J, Snasel V (2013) Spam detection using data compression and signatures. Cybern Syst 44(6–7):533–549
25. Rao GSN (2021) Spam or ham text classification using logistic regression. Turk J Comput Math Educ 12(9):426–433
26. Renuka KD, Visalakshi P (2014) Latent semantic indexing based SVM model for email spam classification
27. Roy PK, Singh JP, Banerjee S (2020) Deep learning to filter SMS Spam. Futur Gener Comput Syst 102:524–533
28. Shehu HA, Browne W, Eisenbarth H (2021) Particle swarm optimization for feature selection in emotion categorization. In: 2021 IEEE Congress on Evolutionary Computation (CEC). IEEE, pp 752–759
29. Sjarif NNA, Azmi NFM, Chuprat S, Sarkan HM, Yahya Y, Sam SM (2019) SMS spam message detection using term frequency-inverse document frequency and random forest algorithm. Procedia Comput Sci 161:509–515
30. Sjarif NNA, YazriwatiYahya S, Azmi NHFM (2020) Support vector machine algorithm for SMS spam classification in the telecommunication industry. Int J Adv Sci Eng Inf Technol 10(2)
31. Torabi ZS, Nadimi-Shahraki MH, Nabiollahi A (2015) Efficient support vector machines for spam detection: a survey. International Journal of Computer Science and Information Security 13(1):11
32. Vieira SM, Mendonça LF, Farinha GJ, Sousa JM (2013) Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. Applied Soft Computing 13(8):3494–3504
33. Wang D, Tan D, Liu L (2018) Particle swarm optimization algorithm: an overview. Soft Comput 22(2):387–408
34. Xia T, Chen X (2020) A discrete hidden Markov model for SMS spam detection. Appl Sci 10(14):5011

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023)
IEEE Xplore Part Number: CFP23AWO-ART; ISBN: 979-8-3503-9663-8

2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192645

Link Prediction in Social Network using Gradient Boosting

Sanket Barman
Dept. of Computer Science and Engineering
Delhi Technological University
Main Bawana Road, New Delhi-110042, India
sanket.barman@gmail.com

Sanyam Srivastava
Dept. of Computer Science and Engineering
Delhi Technological University
Main Bawana Road, New Delhi-110042, India
sanyamsri2001@gmail.com

***Abstract*—Link prediction revolves around network analysis that includes predicting the links that can exist in the near future. Link Prediction has its use in a number of domains such as Social Network Analysis (SNA), bioinformatics, and recommendation systems. In this paper, the use of machine learning classifiers is explored for link prediction and their performance is evaluated against various metrics. After extensive experimentation and evaluation, it is found that Gradient Boosting provides the best results for link prediction. The findings are supported by comparing the performance of different classifiers against similarity scores and ROC curves. The study highlights the importance of using machine learning techniques for link prediction and demonstrates the effectiveness of Gradient Boosting for this task.**

***Index Terms*—Gradient Boost Classifier, Logistic Regression, Link Prediction, Random Forest Classifier, Social Network Analysis**

I. Introduction

In recent years, numerous social network analysis has offered numerous business and research problems like finding influential nodes, community detection, opinion leader detection, fake news, and rumor prediction, link prediction, and many others [1]–[5]. Link prediction is a technique used to analyze a network and in machine learning to check what is the possibility of a future connection between any two nodes in a network. In other words, it predicts whether a link can exist between two nodes that are not connected currently in the network. Fig 1 shows an example of predicting a link between nodes 1 and 3.

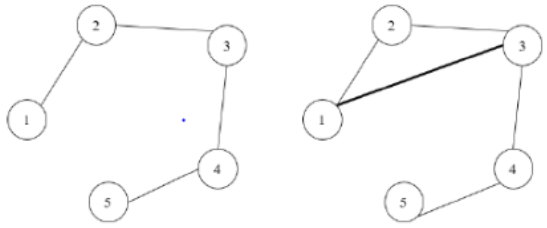


Fig. 1. Link Prediction between nodes 1 and 3

The concept to predict links has grown in popularity ever since it was first introduced in complicated network analysis. In real-life scenario, link prediction has many uses, including forecasting disease outbreaks, predicting link quality in wireless community networks, recommending new friends on social media, planning transportation, making product recommendations for e-commerce applications, and recommending hotels and music, among many others. Link prediction aids in anticipating network development as it expands over time. Therefore, in today's world which is entirely connected, link prediction can be very closely related to information and knowledge architecture.The basis for the link prediction process is the node pair's popularity and closeness. The nodes that are popular will connect with other nodes, and nodes that are adequately similar to one another have a greater likelihood of connecting in the near future, so this relationship appears to be rational. Over the years, numerous link prediction techniques have been proposed in papers. These strategies can be generally categorised into similarity-based, maximum likelihood-based, matrix factorization-based, probabilistic, and random walk methods. Recent research has also looked into the potential of deep learning and machine learning methods. This can be useful for various applications, such as identifying potential friends or collaborators, recommending new connections to users, and detecting fraudulent activities.

Social network analysis (SNA) is an area of study that analyzes social networks and the different people or individuals are related to each other. By analyzing the patterns of these connections, SNA can reveal insights into how information, resources, and influence flow through social networks, and how these networks can impact behavior, attitudes, and outcomes. SNA is applicable in several fields, including sociology, psychology, anthropology, communication studies, management, and marketing, to name a few. SNA uses a range of analytical tools, including centrality measures, clustering algorithms, and network visualization techniques, to analyze social networks and identify key actors, groups, and patterns of interaction. A connection between two nodes is more likely to exist the closer two nodes are to one another. Many methods have been devised for link prediction in social networks, including the Adamic Adar, Jaccard coefficient, preferential attachment methods, etc.

II. Related Works

Over the years, various techniques have been proposed in the literature for predicting links in a network based on

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:50 UTC from IEEE Xplore. Restrictions apply.

the properties of its nodes. These techniques can be broadly categorized into two main types: feature extraction techniques and feature learning techniques [9]. Feature extraction techniques involve manually identifying and extracting relevant features from the network data. These features can include measures such as node degree, common neighbors, and clustering coefficient, among others [10]. In contrast, feature learning techniques involve automatically learning features from the network data using various machine learning algorithms. These techniques can be further categorised into two types: unsupervised and supervised [11].

Based on the idea that nodes frequently create links with other nodes that are similar to them, the most popular approaches for link prediction are those based on similarity. Adamic-Adar (AA) is a similarity measure used in link prediction and community detection in networks [6]. In other words, AA is a measure of how similar the "neighborhoods" of two nodes are. The Jaccard coefficient (JC) is a measure of how similar two sets are [7]. According to its definition, the Jaccard Coefficient is the proportion of distinct neighbors to neighbors in common. Based on the number of neighbors that two nodes share, the Common Neighbors (CN) algorithm forecasts links. The idea is that nodes with a lot of nearby neighbors are more likely to be linked. Based on the idea that nodes with a lot of connections are more likely to draw in new connections, Preferential Attachment (PA) forecasts links. The intuition is that nodes with many connections are more popular, and therefore, more likely to be connected [8]. In an effort to forecast links in different networks, the Resource Allocation algorithm.

In the recent paper regarding node centralities in complex networks, LGBM has been suggested as a better method for predicting links [12]. Another paper uses similarity measures, node centralities, and ML Classifiers to predict links in a social network [13].

III. Proposed Methodology

Using the existing network structure, this research aims to forecast whether a link will exist between a pair of nodes. It is thought that link prediction in complex networks is a problem of binary categorization. The method for comparing different link prediction techniques and combining similarity metrics with machine learning classifiers is described in this section.

To more closely depict the local, semi-local, and global characteristics of the network's nodes, the set of similarity measures was chosen. Different similarity measures have strong correlations, so if machine learning algorithms are applied directly to them, over-fitting is likely to occur. As a result, a particular set of similarity indices is selected that do not exhibit minute correlation and accurately represent the various information present in the nodes. A variety of similarity metrics are used to evaluate the results of the artificial intelligence classifications. The machine learning model uses centrality indices as its features and combines them to generate link predictions. In this paper, Gradient Boosting for link prediction is discussed. In short, the method can be called $GBLP$. Fig. 2 shows a flow diagram of the GBLP algorithm.

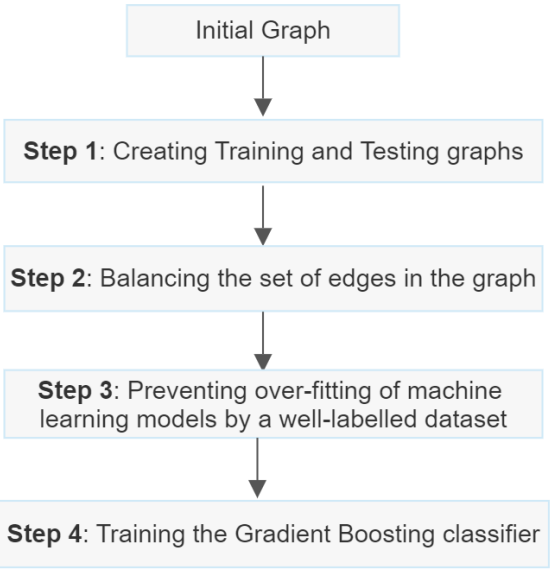


Fig. 2. Flow Diagram of GBLP

The steps involved are as follows:

A. Creating Training and Testing Graphs

In this method, a complex network is represented as a set of vertices V and edges E. A new graph $G_T(V, E_T)$ is created, where E_T is subset of original edge set E, by randomly selecting a subset of edges from the original graph while keeping the same set of vertices. The size of the selected edge set is determined by a splitting ratio applied to the complete graph. The resulting condensed graph, G_T, is then used to train a machine-learning model for link prediction, capable of predicting linkages between nodes in the edge collection $E - E_T$.

By doing so, the link prediction problem can be reduced to predicting the graph $G(V, E')$ where $E' = E - E_T$, using data from the graph $G_T(V, E_T)$. This method is more effective and adaptable to real-world scenarios when compared to other link prediction techniques that train classifiers on the complete network. The training graph, G_T, and testing graph, G', are used to analyze how well the predicted link model works.

B. Balancing the set of edges in the graph

In complicated networks, the number of connected pairs of nodes is very less as compared to the unconnected pairs of nodes, which leads to an unbalanced dataset, causing the machine learning methods to over-fit the negative samples in the dataset. To avoid this, the down-sampling strategy is employed to balance the dataset, where connected node sets are chosen, and the same number of unconnected node pairs are randomly selected. The complement of E_T and E' denotes the group of edges that are absent from the initial network,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:50 UTC from IEEE Xplore. Restrictions apply.

whereas E_U denotes the set of all possible edges in the network. The new graphs, G_T and G', are created using the sampled edges and their complements, respectively.

C. Preventing over-fitting of machine learning models by a well-labeled dataset

In this step, well-labeled datasets are created for the purpose of training and testing, namely D_T and D' respectively. To create these datasets, five similarity measures are chosen, including the Jaccard coefficient, common neighbors, preferential attachment, Adamic Adar, and resource allocation index. These measures are then used to calculate the popularity and similarity measures for the network's nodes and edges. These features are calculated only using the training network's topological data since this is would have access to in a real scenario. The datasets also have an attribute named $label$ that indicates whether or not there are any network connections, and this is determined using the edge's nodes. The value of the label is set to 1 if there is a link between the nodes, and 0 otherwise. Finally, the resulting training and testing datasets are well-labeled and balanced to ensure the machine learning methods do not over-fit on the negative samples.

D. Training the Gradient Boosting Classifier

Well-labeled, evenly distributed datasets DT and D' are created in the final stage to train and test the classifiers. All models are trained in this phase using various machine learning models like logistic regression, random forest, gradient boost, gradient naive Bayes, k nearest neighbor (KNN), and LGBM. Out of all these models, gradient boosting(GB) gives the best result. Thus, the proposed algorithm uses a gradient boosting classifier.

The forecasts of several weak classifiers are combined using a machine learning ensemble technique called a gradient boosting classifier to create a strong classifier. The model is gradually enhanced with decision trees, and each new tree is trained using the errors of the previous one. In other words, it seeks to learn from the model's mistakes in order to improve the precision of predictions. At each step, the algorithm determines the gradients of the loss function in relation to the anticipated values. Then, a fresh decision tree is fitted using the negative gradient of the loss function. The forecasts of the new tree and the previous trees are merged using a weighted sum after the new tree has been properly incorporated into the model. Based on how well each tree did on the training set of data, a weight was assigned to it; trees that performed poorly were given a lower weight. This allows the model to focus on the regions of the feature field that are challenging to classify and lowers overall model error. The final model is then used to generate forecasts on new data by individually processing the data through each decision tree and combining their outputs using the same weighted sum.

IV. Datasets and Evaluation metrics

A. Datasets

This section includes various datasets and the metrics that have been used for evaluation in the paper:

1) Caltech36 [14]: This dataset is basically a Facebook buddy graph made up of Caltech, Pasadena students. An individual is represented by each node of the graph, while friendships or other social connections are represented by the network linkages. It has approximately 16656 edges and 769 nodes. It can be found at https://networkrepository.com/socfb-Caltech36.php.
2) PolBlogs [15]: The PolBlogs network, where the nodes can be considered as pages of a blog and the edges can be considered as hyperlinks between those pages of a blog. It depicts the interplay between US political blog pages. It has 19021 connections and 1222 nodes.
3) DBLP [16] : It serves as a reference network for DBLP bibliography of computer science. In this, each node stands in for a piece of writing, and each edge for a reference made between those pieces of writing. There are 49743 edges and 12591 nodes. It is availaible at https://snap.stanford.edu/data/ego-Facebook.html.
4) Fb-Tv Show [17]: This is a social network data of Facebook which displays the common likes between various TV program sites on Facebook. The network's nodes stand in for the TV program pages, while its edges stand in for shared likes between the pages. There are around 17260 edges and 3892 nodes. It can be found at https://networkrepository.com/fb-pages-tvshow.php.
5) Delaunay-N12 [18] : This network dataset is in the category of DIMACS10.These files were created by triangulating random locations in the unit square using the Delaunay method. It consists of 4100 nodes and 12300 edges. It is availaible at https://networkrepository.com/delaunay-n12.php.

TABLE I
Statistics of Dataset

Dataset	Nodes	Edges
CalTech36	769	16656
PolBlogs	1222	19021
DBLP	12591	49753
Fb-Tv Show	3892	17260
Delaunay- N12	4100	12300

B. Evaluation Metrics

AUC (Area under the curve), recall, precision and also the F1 score are some of the traditional assessment metrics employed to objectively assess how effective the suggested algorithm is. The performance of the system is compared to a number of other baseline and existing link prediction techniques. The predictive performance of different link prediction

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:50 UTC from IEEE Xplore. Restrictions apply.

algorithms is clearly shown by all the evaluation measures. In the following equations, TN is the amount of samples which are negative but have been correctly predicted while TP is the amount of samples which are positive and are predicted correctly. FP is the amount of actual negative samples which have been predicted as positive while FN is the amount of actual positive samples which have been predicted as negative.

1) AUC: The area beneath the AUC curve is used to evaluate the performance of a machine learning model. AUC's value has a range from 0.0 to 1.0. The algorithmic accuracy increases with increasing AUC value. If there are T independent comparisions, T' links which are missing with a high score and T'' links which are missing and doesn't exist, with a same score, then AUC is written as in Eq. 1:

$$AUC = \frac{T' + 0.5T''}{T} \tag{1}$$

2) Accuracy: Accuracy is the percentage of accurate forecasts made by the model. Accuracy can also be determined in terms of positives and negatives for binary categorization, as shown in 2:

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{2}$$

3) Precision: Precision is the percentage of correctly classified positive classifications.The precision score is calculated in 3

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

4) Recall: Recall is an indicator of performance that seeks to indicate what proportion of genuine positives are accurately detected. It is defined in 4:

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

5) F1 score: The F1 score basically represents the balance in recall and precision balance. It can be stated as the harmonic mean of recall and precision. as shown in Eq. 5:

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

V. Results and Analysis

Here, using a variety of evaluation metrics and the ROC Curve, the performance of the suggested algorithm and other algorithms is evaluated on each dataset.

TABLE II
COMPARISION OF DIFFERENT MACHINE LEARNING MODELS FOR LINK PREDICTION BASED ON AUC, ACCURACY, PRECISION, RECALL AND F1 SCORE FOR CALTECH36 DATASET

Models	AUC	Accuracy	Precision	Recall	F1 Score
RF	0.727	0.728	0.708	0.714	0.711
LGBM	**0.800**	0.746	0.723	0.745	0.733
LR	0.764	0.53	0.485	**1.00**	0.653
KNN	0.637	0.638	0.615	0.613	0.614
Gausian NB	0.683	0.694	**0.762**	0.506	0.608
GBLP	0.796	**0.760**	0.751	0.732	**0.741**

TABLE III
COMPARISION OF DIFFERENT MACHINE LEARNING MODELS FOR LINK PREDICTION BASED ON AUC, ACCURACY, PRECISION, RECALL AND F1 SCORE FOR POLBLOGS DATASET

Models	AUC	Accuracy	Precision	Recall	F1 Score
RF	0.763	0.762	0.797	0.733	0.763
LGBM	0.820	0.75	0.799	0.698	0.745
LR	0.821	0.476	0.0	0.0	0.0
KNN	0.718	0.72	0.724	**0.752**	0.738
Gausian NB	0.711	0.702	**0.874**	0.504	0.639
GBLP	**0.829**	**0.774**	0.817	0.733	**0.773**

TABLE IV
COMPARISION OF DIFFERENT MACHINE LEARNING MODELS FOR LINK PREDICTION BASED ON AUC, ACCURACY, PRECISION, RECALL AND F1 SCORE FOR DBLP DATASET

Models	AUC	Accuracy	Precision	Recall	F1 Score
RF	0.773	0.774	**0.766**	0.756	0.761
LGBM	0.860	**0.800**	0.761	0.845	**0.801**
LR	0.838	0.48	0.478	**1.00**	0.646
KNN	0.742	0.742	0.722	0.744	0.733
Gausian NB	0.580	0.562	0.521	0.975	0.679
GBLP	**0.867**	0.784	0.734	0.857	0.791

TABLE V
COMPARISION OF DIFFERENT MACHINE LEARNING MODELS FOR LINK PREDICTION BASED ON AUC, ACCURACY, PRECISION, RECALL AND F1 SCORE FOR FBTVSHOW DATASET

Models	AUC	Accuracy	Precision	Recall	F1 Score
RF	0.907	0.908	0.926	**0.880**	0.903
LGBM	0.964	0.904	0.937	0.860	0.897
LR	0.931	0.514	0.0	0.0	0.0
KNN	0.895	0.896	0.910	0.872	0.891
Gausian NB	0.897	0.898	0.925	0.860	0.891
GBLP	**0.965**	**0.922**	**0.959**	0.877	**0.916**

TABLE VI
COMPARISION OF DIFFERENT MACHINE LEARNING MODELS FOR LINK PREDICTION BASED ON AUC, ACCURACY, PRECISION, RECALL AND F1 SCORE FOR DELAUNAY-N12 DATASET

Models	AUC	Accuracy	Precision	Recall	F1 Score
RF	0.796	0.796	**0.846**	0.724	0.780
LGBM	**0.831**	0.802	0.832	0.756	0.792
LR	0.824	0.811	0.833	**0.780**	0.805
KNN	0.686	0.686	0.644	0.752	0.705
Gausian NB	0.812	**0.812**	0.833	**0.780**	**0.806**
GBLP	0.825	0.808	0.835	0.768	0.800

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:50 UTC from IEEE Xplore. Restrictions apply.

A. Performance Comparision

The computed results for 5 datasets is shown below. It is compared on the basis of 5 evaluation metrics in 6 different models.

The analysis of the findings indicates that GBLP is the best model for link prediction in most of the datasets that were examined. $TableII$ presents a comparison of the performance for Caltech36 dataset. It shows that the proposed method achieves the highest accuracy. Furthermore, in $TableIII$, the performance of polBlogs dataset is presented, which shows that the GBLP method outperforms other methods not only in terms of AUC and accuracy but also in its F1 score, making it the most suitable model for this dataset. Regarding the DBLP dataset in $TableIV$, since the difference in accuracy between LGBM and GBLP is negligible, both the methods can be considered as viable options. However, for the FbTVShow dataset in $TableV$, the GBLP method shows the best results in most of the metrics, indicating that it is the most effective model for this dataset. Finally, in $TableVI$, although the values of the proposed method are not the highest, the difference between the metrics of GBLP and other methods is negligible. Therefore, it can still be considered a viable option. Overall, the study highlights the effectiveness of Gradient Boosting.

B. ROC Curve

Below are the ROC curves of various algorithms on different datasets. The "False Positive Rate" is represented by the x axis, while the "True Positive Rate" is represented by the y axis. The statistical test used to compare two ROC Curves involves calculating the area under each ROC curve (AUC) and then comparing the two AUC values to determine significant difference between the two curves. These curves clearly represent that the gradient boosting(GB) performs best among all the models. The AUC scores of each model is also mentioned in the plots of each dataset.

Based on the evaluation results, it can be observe that the overall accuracy of a model has a direct impact on the shape and position of its ROC curve. Specifically, a model with higher accuracy tends to have a ROC curve that is closer to the top left corner, which signifies better performance in distinguishing positive and negative instances. In $Fig.3$, the ROC curves of LGBM and GBLP are very close to y axis, indicating that it has better accuracy. The $Fig.4$ and $Fig.5$ for PolBlogs and DBLP datasets, the curve for GBLP is closer to the top left corner, again indicating better accuracy compared to other models. A similar trend can be observed in the $Fig.6$ for FbTVShow dataset, where the GB model has better accuracy. There is a negligible difference in accuracy between the GBLP and LGBM model in $Fig.7$. So, the majority of evaluation results suggest that the GBLP model performs well and can be considered a viable method for link prediction.

VI. Conclusion

This paper focuses on how different machine learning algorithms can be used to predict future links in a network and highlights the effectiveness of Gradient Boosting for this

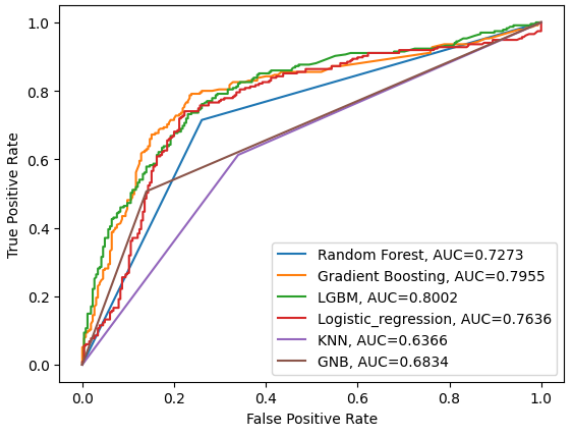


Fig. 3. ROC Curve obtained on Caltech36 dataset for different algorithms

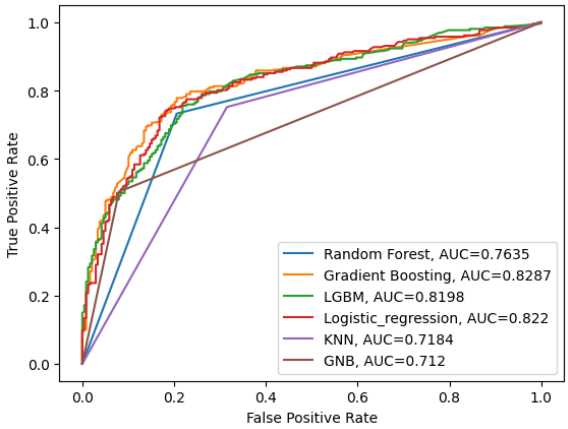


Fig. 4. ROC Curve obtained on PolBlogs dataset for different algorithms

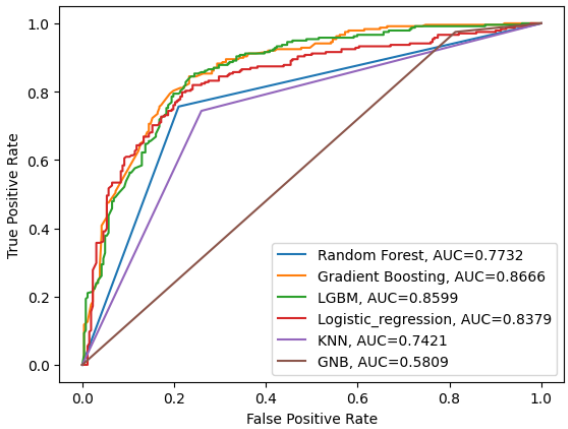


Fig. 5. ROC Curve obtained on DBLP dataset for different algorithms

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:50 UTC from IEEE Xplore. Restrictions apply.

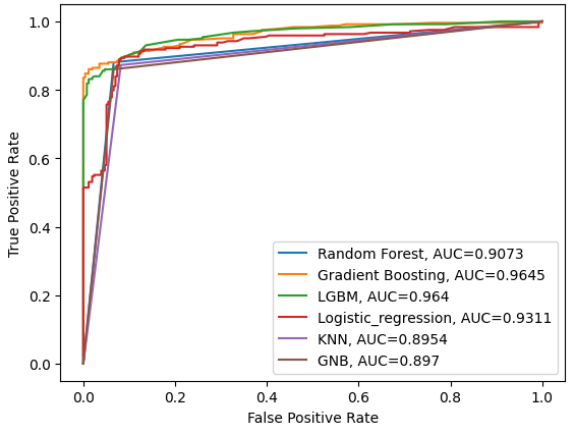


Fig. 6. ROC Curve obtained on FbTVShow dataset for different algorithms

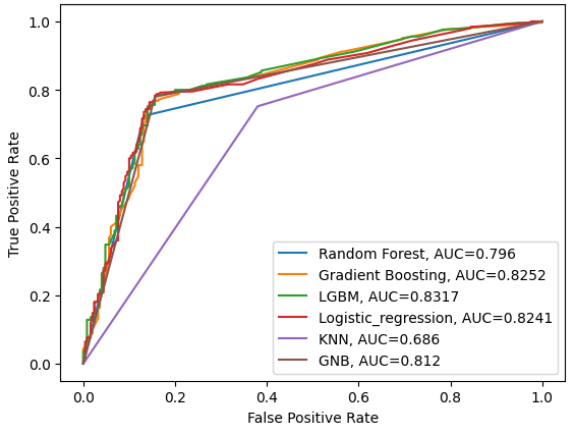


Fig. 7. ROC Curve obtained on Delaunay-N12 dataset for different algorithms

topic. The high accuracy of GBLP shows its ability to handle large datasets and its capacity to handle complex relationships between the variables. To evaluate the performance of the machine learning models, it is important to consider the performance metrics beyond accuracy. In some cases of comparision between LGBM and GBLP models, it is significant to observe that though there is a very minute difference in the accuracy, GBLP outperforms LGBM in other important performance metrics. Thus, it is important to analyze on the basis of other performance metrics too, not just accuracy, while evaluating the machine learning models and consider a broader range of performance metrics. Overall, the performance tables shows that GBLP is a strong performer for link prediction, while LGBM can be a viable alternative with comparable performance. These models are quite important in predicting links, where accuracy and efficiency are critical factors to be considered for success of an algorithm. After analysis of ROC curves, it is observed that it is important to consider accuracy of a binary classification model while evaluating its performance, and GBLP model is an optimal choice for link prediction tasks.

REFERENCES

[1] S. Kumar, A. Kumar, B.S. Panda, "Identifying influential nodes for smart enterprises using community structure with Integrated Feature Ranking". IEEE Transactions on Industrial Informatics, 19(1), 703-711, 2022,
[2] S. Kumar, A. Mallik, B. S. Panda, "Influence maximization in social networks using transfer learning via graph-based LSTM". Expert Systems with Applications, 212, 118770, 2023
[3] S. Kumar, A. Mallik, S. S. Sengar, "Community detection in complex networks using stacked autoencoders and crow search algorithm". The Journal of Supercomputing, 79(3), 3329-3356, 2023
[4] S. Kumar, A. Kumar, A. Mallik, S. Dhall," Opinion leader detection in Asian social networks using modified spider monkey optimization". Transactions on Asian and Low-Resource Language Information Processing, 2022
[5] S. Kumar, A. Kumar, A. Mallik, R. R. Singh, "OptNet-Fake: Fake News Detection in Socio-Cyber Platforms Using Grasshopper Optimization and Deep Neural Network". IEEE Transactions on Computational Social Systems, 2023
[6] L.A. Adamic, E. Adar, "Friends and neighbors on the Web". Social Networks 25(3), 211–30 (2003)
[7] P. Jaccard, "Etude comparative de la distribution florale dans une portion des Alpes et des Jura". Bull Soc Vaudoise Sci Nat 37, 547–79 (1901)
[8] A.L. Barabasi, R. Albert, "Emergence of scaling in random networks". Science 286(5439), 509–12 (1999)
[9] Ece C. Mutlu, Toktam A. Oghaz, Amirarsalan Rajabi, Ivan Garibay, "Review on graph feature learning and feature extraction techniques for link prediction". arXiv:1901.03425 (2019)
[10] N.M.A. Ibrahim, L. Chen, "Link prediction in dynamic social networks by integrating different types of information". Applied Intelligence 42(4), 738–750 (2015)
[11] B. Pandey, P.K. Bhanodia, A. Khamparia, D.K. Pandey, "A comprehensive survey of edge prediction in social networks: techniques, parameters and challenges". Expert Systems with Applications 124, 164–81 (2019)
[12] S. Kumar,A. Mallik, B.S. Panda, "Link prediction in complex networks using node centrality and light gradient boosting machine". World Wide Web 25, 2487–2513 (2022)
[13] S. Anand, Rahul, A. Mallik, S. Kumar, "Integrating node centralities, similarity measures,and machine learning classifiers for link prediction". Multimedia Tools and Applications 81(27), 38593–38621 (2022)
[14] E. Kelsic, P. Mucha, M. Porter, "Comparing community structure to characteristics in online collegiate social networks". SIAM Review 53,10, 1137/080734315 (2008)
[15] L.A. Adamic, N. Glance, "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog". Proceedings of the 3rd International workshop on link discovery ,36-43 (2005)
[16] R.A. Rossi, N.K. Ahmed, "The network data repository with interactive graph analytics and visualization". Proceedings of the AAAI Conference on Artificial Intelligence 29, 1 (2015)
[17] B. Rozemberczki, R. Davies, R. Sarkar, C. Sutton, "GEMSEC: graph embedding with self clustering". Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, 65–72 (2019)
[18] R.A. Rossi, N.K. Ahmed, "The Network Data Repository with Interactive Graph Analytics and Visualization". Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 4292–4293 (2015)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:32:50 UTC from IEEE Xplore. Restrictions apply.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Low Supply QFGMOS CCCII with Low-pass and High-pass Filter Applications

Ujjval Gupta
Electrical Engineering
Delhi Technological University
Rohini, Delhi
ujjvalgupta70@gmail.com

Subrut Shanker Yagnik
Electrical Engineering
Delhi Technological University
Rohini, Delhi
subrutyagnik17@gmail.com

Vaibhav Gangian
Electrical Engineering
Delhi Technological University
Rohini, Delhi
vaibhavgangian2000@gmail.com

Anupama
Electrical Engineering
Delhi Technological University
Rohini, Delhi
anupamaverma.91@gmail.com

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205701

***Abstract*—In this paper, a QFGMOS based current controlled current conveyor (CCCII) is proposed, which decreases the required voltage to ±0.5V and the power consumption. It improves the linearity of the current characteristics, and also reduces the parasitic resistance than basic CCCII. The proposed circuit is also used in low-pass and high-pass filter circuits, whose performance is obtained in PSPICE using TSMC 180 nm CMOS technology. The results obtained for the low-pass filter and high pass filter show an improvement in the bandwidth of 1.579 times and 1.294 times respectively by using QFGMOS in CCCII. Also, the current linearity in the proposed QFGMOS CCCII has enhanced by twice.**

Keywords— bandwidth, current conveyor, filters, low voltage, quasi floating gate MOSFET (QFGMOS)

I. INTRODUCTION

One of the major drivers of electronic circuit design is the desire and the need for power reduction. To compete in a market that is changing quickly, it is necessary to be able to create less expensive, higher performing, and low power gadgets. Nowadays, the need to reduce the consumed power has further extended to applications such as home entertainment and computer farms from battery-powered applications like remote controls and portable devices. Innovations in circuit designing are heavily driven by large data centers due to their requirement for low power consumption [1].

The past decade has seen a significant rise in the creation of new low power CMOS circuit design, largely due to the widespread use of portable electronic gadgets such as tablets, mobile phones, and laptops. Several methods and approaches have been explored for creating low power with less supply voltage circuits, one of the significant is the current-mode circuit approach. The current-mode circuits has numerous benefits as compared to the voltage-mode, including higher speed, greater bandwidth, simpler circuitry, better operational accuracy and lower consumption of power. In addition, current-mode circuits can improve performance in terms of bandwidth, low-voltage, and slew rate characteristics. A key benefit of the current-mode technique is to overcome the constant gain-bandwidth product limitation that exists in voltage-mode techniques [4].

The current conveyor circuit is a key and practical component of common mode applications. The first generation of the current conveyor which is known as CCI, was first introduced by Sedra and Smith in 1968 [2]. In CCI, the input impedance of the ports, X as well as Y, is zero to sink the currents. This affects the flexibility of the CCI, which reduces its applications. So, Sedra and Smith subsequently introduced the second generation of the current conveyor (CCII) in 1970 [3]. The CCII has both a high impedance and a low impedance input ports to overcome the limitations of the CCI. Due to the good performance and functional versatility of the CCII, it has become a viable alternative to conventional op-amps. This has led to its extensive use in high-performance electronic components such as oscillators, differentiators, integrators, and filters that operate in voltage or current mode [5–9].

Despite the CCII having several advantages over CCI, it is unable to limit the parasitic resistance (R_x) that exists at port X. To overcome this issue, a second-generation current controlled current conveyor (CCCII), which has the benefit of electronic tunability over the CCII [12], was developed. The CCCII is capable of controlling the parasitic resistance, and hence provides electronic tunability. In accordance with the bias current, the parasitic resistance (R_x) at port X can be altered which can be used to change the quality factor or cutoff frequency for applications like filters [17]. The CCCII further does not need any external passive resistors, which shrinks the chip area and ensures the consumption of lesser power.

To obtain circuits that require low voltage and circuits that consume less power, the main requirement is to reduce the size of transistors. In order to ensure the reliability of the device, as the size of the transistor decreases, the supply voltage must also decrease. However, the threshold voltage (V_{TH}) does not decrease in the same way because of its large impact on leakage current and noise margin. Among the various techniques suggested to address this issue, QFGMOS is an important technology for making low power and low voltage circuits [11,13-14].

This paper proposes quasi-floating MOSFETs (QFGMOS) to improve the characteristics of CCCII while maintaining low power dissipation. A low supply of ±0.5 V is used to implement the proposed QFGMOS based CCCII,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:16:53 UTC from IEEE Xplore. Restrictions apply.

resulting in a reduced threshold voltage (V_{TH}), improved transconductance and linearity of the MOSFET. Also, the improvement of bandwidth is portrayed in both filters, which use two QFMOS based CCCIIs with two capacitors. Without modifying the circuit structure, both the low-pass and the high-pass filters can be realised by choosing the appropriate filter input signal.

II. Design of CMOS Based BASIC CCCII

A. Working Principle of CCCII

Three terminals or ports X, Y, and Z are present on the CCCII. A voltage follower is present between X and Y terminals, and a current mirror is present between X and the Z terminals. As a result, X port will have the same voltage as Y port and X port will have the same current as Z port.

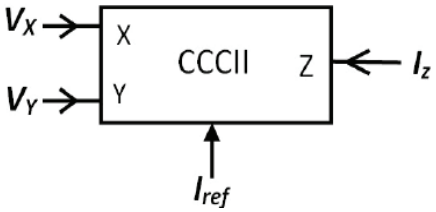


Fig. 1. Symbolic representation of CCCII [18]

The CCCII symbol diagram is given in Fig. 1. In this, Y represents for the input voltage port with a high input impedance, X for the voltage port, Z for the output current port and I_{ref} for the reference current.

The characteristic of the inverted or non-inverted CCCII is given by the equation 1 [10]-

$$\begin{bmatrix} I_Y \\ V_X \\ I_Z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & R_X & 0 \\ 1 & \pm 1 & 0 \end{bmatrix} \quad (1)$$

Where V_x, V_y, V_z and I_x, I_y, I_z are the voltage and current of ports X, Y and Z respectively. Fig. 2 represents the parasitic resistance (R_x) in CCCII. The current through port X (I_x) is given by the equation 2.

$$I_x = I_z \pm \frac{V_x - V_y}{R_x} \quad (2)$$

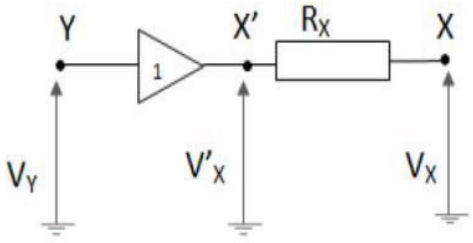


Fig. 2. Parasitic resistance (Rx) in CCCII [18]

B. Circuit Diagram of CMOS based basic CCCII

Fig. 3 represents the basic CCCII model. It contains two main parts, the negative part and the positive part. The transistors from M_1 to M_5 represent the positive part and the negative part of the model is represented by transistors from M_6 to M_{10}.

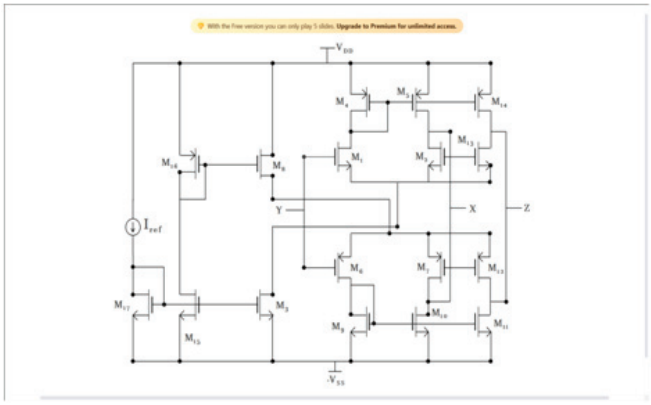


Fig. 3. Basic CCCII model [18]

The differential pair of positive inputs X and Y is designed by the M_1 and M_2 transistors. The drain and gate terminals of the transistor M_2 provide a low impedance to port X. An active load is provided by the transistors M_4 and M_5. The current biasing for the differential pair is provided by M_3 [18].

Similar to the positive part, in the negative part a differential pair is formed because of the M_6 and M_7 transistors. An active load is formed using transistors M_9 and M_{10}. Transistor M_8 is responsible for behaving as the current biasing transistor.

The transistors M_{11}-M_{14} represent a current mirror between ports X and Z. Another current mirror is formed by transistors M_{15}-M_{17} which mirror the reference current to transistors M_8, and M_3. The parasitic resistance is expressed in the following equation 3 [18],

$$r_x = \frac{1}{g_{m2}} \,||\, \frac{1}{g_{m7}} \quad (3)$$

III. Proposed QFGMOS Based CCCII

This section describes the quasi-floating gate technique and the circuit that has been proposed. This circuit is then further used in the filter application to obtain the bandwidth of the low pass and the high pass filters.

A. Quasi Floating Gate Technique

In recent years, Floating Gate MOSFET (FGMOS) and Quasi-Floating Gate MOSFET (QFGMOS) technologies are being employed in circuits to improve their performance as well as to ensure low voltage and low power applications. These techniques are widely employed in circuits like op-amps, OTAs (Operational trans-conductance amplifiers) current mirrors, multipliers etc. [13-15].

The higher threshold voltage of a MOSFET creates issues in low voltage operation because of which newer and better techniques were needed to be developed [15]. In response to this problem, FGMOS was created whose threshold voltage can be varied and can be reduced by applying an appropriate bias voltage through a large capacitor to one of the input terminals of the MOSFET. However, the presence of a large valued capacitance increases the silicon area and reduces the gain bandwidth product which decreases the transconductance. Also, during fabrication, a charge of a significant amount can be trapped by an FGMOS which can cause DC offsets [13].

To overcome the shortcomings of the FGMOS, the QFGMOS was developed. Like the FGMOS, in QFGMOS

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:16:53 UTC from IEEE Xplore. Restrictions apply.

there is a capacitance between the input terminals and the gate terminals of the MOSFET and a very high resistance is used to establish a weak connection between the gate terminal and one of the supply rail terminals. The large resistance is usually implemented by using a diode-connected reverse-biased transistor. Due to the usage of this large resistance, the gate dc voltage is set to a power rail and the initial charge problems are prevented, along with this the supply voltage requirements also diminish [14].

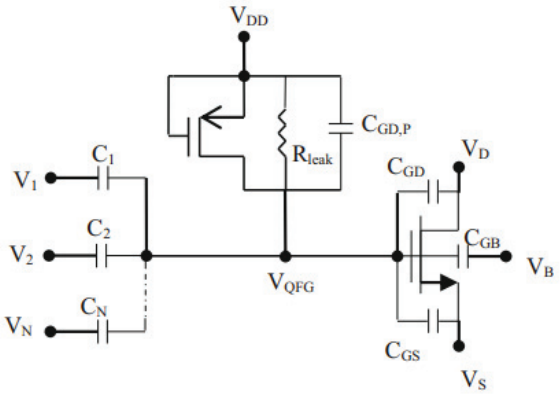


Fig. 4. Equivalent circuit of QFGMOS

In Fig. 4 the C_{GD} and R_{leak} represent the equivalent parasitic resistance and the equivalent resistances. Between the control and the floating gate, there are 'n' capacitances, and to each of the n capacitances, a supply voltage is connected. C_{GD} represents the capacitance that is present between the gate terminal and the drain terminal, C_{GB} represents the capacitance that is present between the gate and the body and C_{GS} is the capacitance that is present between the gate and the source. The voltage that is present at the floating gate is expressed by the following equation 4 [11].

$$V_{QFG} = \frac{sR_{leak}}{1 + sR_{leak}C_T}\left(\sum_{k=1}^{n} C_K V_K + C_{GS} + C_{GD}V_D + C_{GB}V_B\right) \quad (4)$$

B. Circuit Diagram of Proposed QFGMOS Based CCCII

Fig. 5 represents the proposed QFGMOS based CCCII. The transistors M_1, M_2, M_6 and M_7 have been connected with transistors M_{18}, M_{19}, M_{20} and M_{21} to create a QFGMOS configuration of the current controlled current conveyor.

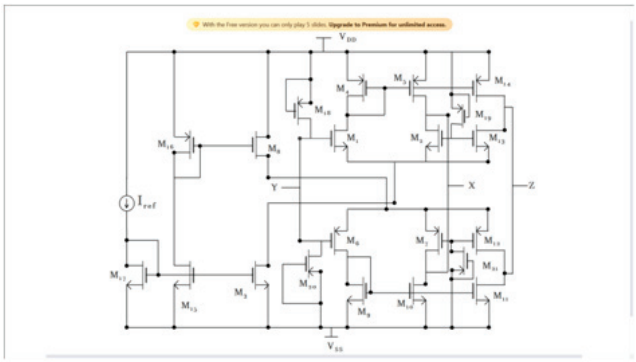


Fig. 5. Proposed QFGMOS CCCII

C. Application as a Filter

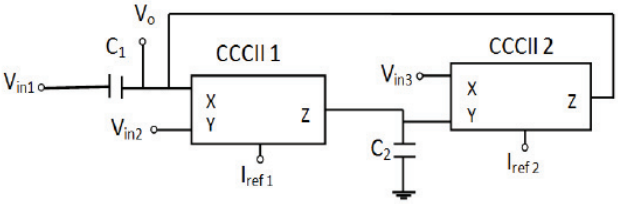


Fig. 6. Filter configuration using two proposed QFGMOS CCCII [5]

In Fig. 6, two-CCCII and capacitor-based configuration to represent filters has been proposed by Jaikla [16]. In the present work, the same circuit has been implemented using two proposed QFGMOS CCCII. The filter transfer equation is given by equation 5 [16].

$$V_o = \frac{s^2 C_1 C_2 R_{x1} R_{x2} V_{in2} + V_{in2}(sC_2R_{x2} + 1) - sC_1R_{x1}V_{in3}}{s^2 C_1 C_2 R_{x1} R_{x2} + sC_2R_{x2} + 1} \quad (5)$$

By altering the values of V_{in1}, V_{in2} and V_{in3}, the transfer functions for high-pass and low-pass filters can be attained.

Table I represents the value of input voltages for filters used in Fig. 6 for the operation of the filter as a low-pass or a high-pass filter.

TABLE I. INPUT VOLTAGES FOR FILTERS

Filter Response	Input Selection		
V_o	V_{in1}	V_{in2}	V_{in3}
Low-Pass	0	1	1
High-Pass	1	0	0

IV. SIMULATION RESULTS

The dimensions of the various transistors that have been used in this paper are given in Table II.

TABLE II. MOSFET DIMENSIONS

Transistors	W(μm)/L(μm)
M_3, M_4, M_{11}	4.5/0.18
M_7, M_8, M_9, M_{15}	5/0.18
M_1, M_2, M_{10}	17.5/0.18
M_5, M_6, M_{12}, M_{16}, M_{17}	15/0.18
M_{13}	10/0.18
M_{14}	27.5/0.18
M_{18}, M_{19}, M_{20}, M_{21}	5/0.18

The analysis of the proposed QFGMOS based CCCII has been performed on the PSPICE technology using the 180 nm CMOS TSMC Technology. For the analysis, the supply voltage V_{DD} and V_{SS} are fixed at ±0.5V, I_{ref} is taken as 200μA and the capacitances used have been taken as 100pF.

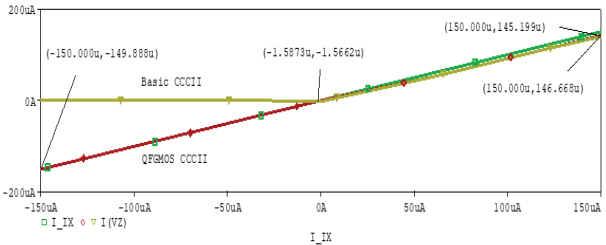


Fig. 7. Current characteristics of basic and proposed QFGMOS CCCII

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:16:53 UTC from IEEE Xplore. Restrictions apply.

The current characteristics are shown in Fig. 7. From the graph, it is observed that the basic CCCII is capable of mirroring the current from -1.587 µA to 145.2 µA, whereas, the proposed QFGMOS CCCII is capable of mirroring the current from -149.888 µA to 146.668 µA. This shows that the proposed QFGMOS CCCII improves linearity of current characteristic by approximately twice.

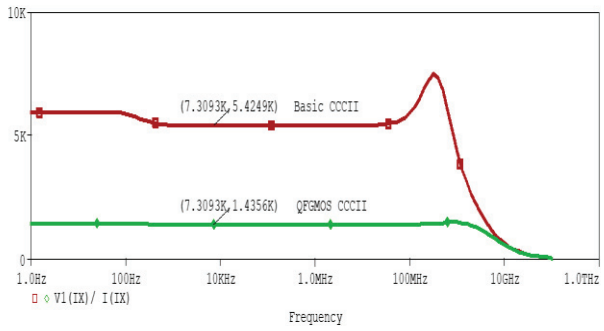


Fig. 8. Frequency response of the parasitic resistance for basic and proposed QFGMOS CCCII

The frequency response for the basic CCCII and the proposed QFGMOS CCCII is depicted in Fig. 8. It indicates that the parasitic resistance reduces from 5.4249 kΩ (basic CCCII) to 1.4356 kΩ (proposed QFGMOS CCCII).

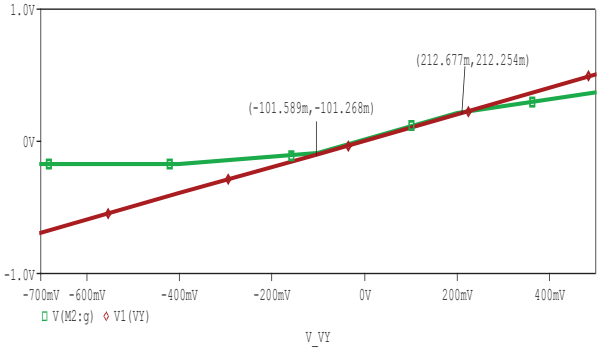


Fig. 9. Voltage characteristics of proposed QFGMOS based CCCII

The voltage characteristics of the proposed QFGMOS CCCII is given in Fig. 9. It shows that V_X and V_Y are nearly same in the region from -101.589 mV to 212.677 mV for a reference current of 200 µA.

The proposed QFGMOS CCCII and basic CCCII are compared and summarized in Table III. The comparison parameters indicate the proposed QFGMOS CCCII requires less power than basic CCCII at same power supply.

TABLE III. BASIC CCCII AND PROPOSED QFGMOS CCCII COMPARISON

Specifications	Basic CCCII [18]	Proposed QFGMOS CCCII
Type	MOS	MOS
Supply Voltage	± 0.5V	± 0.5V
Number of Transistors	17	21
Max Power Consumption (mW)	0.242	0.228
Parasitic Resistance (KΩ)	5.4249 kΩ	1.4356 kΩ
Length of transistor	0.18µ	0.18µ

Low pass and high pass filters are simulated using basic CCCII and proposed QFGMOS CCCII. Fig. 10 depicts the simulation of low-pass filter using basic CCCII and proposed QFGMOS CCCII in which the values of V_{in1}, V_{in2} and V_{in3} are taken as 0, 1 and 1 in accordance with Table I.

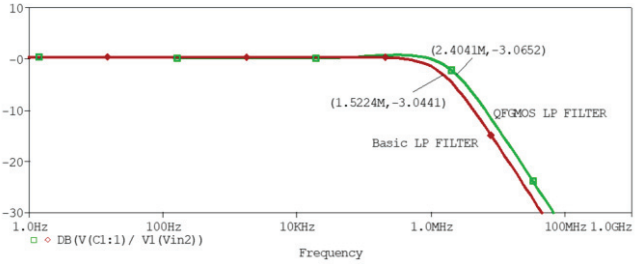


Fig. 10. Low-pass filters of basic and proposed QFGMOS CCCII

From Fig. 10, it is observed that the bandwidth for the basic CCCII based low-pass filter is 1.5224 MHz and the bandwidth for the proposed QFGMOS CCCII based low-pass filter is 2.4041 MHZ. This indicates an increase in bandwidth for the proposed QFGMOS CCCII based low-pass filter by 1.579 times that of the basic CCCII based low-pass filter.

Fig. 11 represents the simulation of high-pass filter using basic CCCII and proposed QFGMOS CCCII in which, the values of V_{in1}, V_{in2} and V_{in3} are taken as 1, 0 and 0 in accordance with Table I. It is observed that the bandwidth for the basic CCCII based high-pass filter is 981.748 KHz and the bandwidth for the proposed QFGMOS CCCII based high-pass filter is 1.2706 MHz. This indicates an increase in bandwidth of the proposed QFGMOS CCCII based high-pass filter by 1.294 times to that of the basic CCCII based high-pass filter.

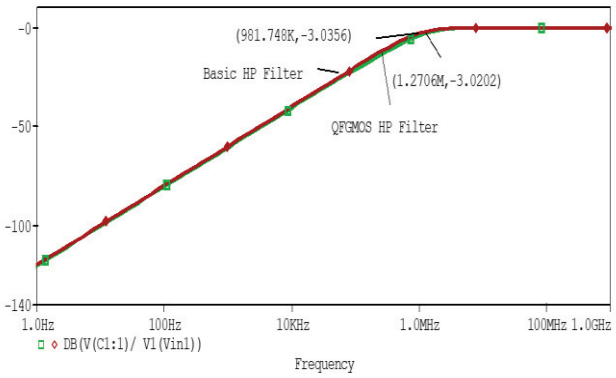


Fig. 11. High-pass filters of basic and proposed QFGMOS CCCII

Table IV shows the basic CCCII and proposed QFGMOS CCCII filters comparison.

TABLE IV. BASIC CCCII AND PROPOSED QFGMOS CCCII FILTERS COMPARISON

S.NO.	Filter Type	Basic CCCII Filter	Proposed QFGMOS CCCII Filter	Enhancement
1	Low Pass	1.5224 MHz	2.4041 MHz	1.579 times
2	High Pass	981.748 KHz	1.2706 MHz	1.294 times

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:16:53 UTC from IEEE Xplore. Restrictions apply.

V. Conclusion

In this paper, QFGMOS based CCCII is proposed whose validity is verified by plotting the current and voltage characteristics. The result shows that the proposed QFGMOS CCCII improves the current linearity by maintaining low power consumption than basic CCCII. Also, the proposed circuit is used to implement the high pass filter and low pass filter which are compared to basic CCCII based filters. It is observed that the bandwidth for proposed QFGMOS CCCII based filters is much higher than the bandwidth of basic CCCII based filters. The increase in bandwidth ranges from 1.294 to 1.579 times for the same reference current (I_{ref}). As the proposed circuit uses very low voltages of ±0.5 V and is hence apt for various low power and low voltage applications.

References

[1] Coro García Alberdi, Low Power Channel Selection Filtering for Highly Integrated Wireless Receivers, 2013.

[2] A. S. Sedra, K. C. Smith, "The Current Conveyor – A New Circuit Building Block," Proceeding Circuits and Systems, vol. 56, pp. 1368–1369, 1968.

[3] K. C. Smith, A. S. Sedra "A Second-Generation Current Conveyor and its Applications," IEEE Transactions on Circuit Theory, vol. 17, pp. 132–134, 1970.

[4] G. Ferri and N. Guerrini, "Low voltage low power CMOS current conveyors," Kluwer Academic Publisher, Boston, 2003

[5] J.-W. Hrong, "High input impedance voltage mode universal biquad filters with three inputs using plus type CCII," Int. J. Elect., vol. 91, pp. 465- 475, 2004.

[6] C.M. Chang, C.L. Hou, J.-W. Horng, W.U. Chung, H.W. Tang, Y.H. Wen, "Quadrature oscillators using CCIIs," Int. J. Elect., vol. 92, pp. 21-31, 2005.

[7] K.K. Dey, A.A. Khan, S. Bimal, S.S. Roy, "Novel RC sinusoidal oscillator using second generation current conveyor," IEEE Trans. On Instrumentation and Measurement, vol. 54, pp. 2402-2406, 2005.

[8] J. Sabadell, P.A. Martinez, C. Aldea, S. Celma, "Variable frequency sinusoidal oscillator based on CCII+," IEEE Trans. Circuits Systems-1: Fundamental theory and applications, vol. 46, pp. 1386-1390, 1999.

[9] G. Ferri, N. Guerrini, E. Silverii, A. Tatone, "Vibration Damping Using CCII-Based Inductance Simulators Instrumentation and Measurement," IEEE Transactions on Digital Object Identifier 10.1109/TIM.2007.913762, 57(5), pp. 907-914,2008.

[10] Q. Zhang, C. Wang, J. Sun, and S. Du, "A new type of current conveyor and its application in fully balanced differential current mode elliptic filter design," Journal of Electrical Engineering, vol. 62, no. 3, pp. 126-133, 2011.

[11] J. M. A. Miguel, A. J. Lopez-Martin, L. Acosta, J. Ramirez-Angulo and R. G. Carvajal, "Using Floating Gate and Quasi-Floating Gate Techniques for Rail-to-Rail Tunable CMOS Transconductor Design," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 7, pp. 1604-1614, July 2011.

[12] B. Parveen, N. A. Shah, S. Z. Iqbal, "SITO high output impedance transadmittance filter using FTFNs ," Analog Integrated Circuits and Signal Processing, vol. 40, no. 1, pp. 87- 89, 2004.

[13] J. R. Angulo, A. J. Lopez-Martin, R. G. Carvajal and F. M. Chavero, "Very low voltage analog signal processing based on quasi floating gate transistors," IEEE Journal of Solid State Circuits, vol. 39, pp. 434-442, 2004.

[14] R. Gupta, S. Sharma, S.S. Jamuar, A low voltage current mirror based on quasi-floating gate mosfets, 2010 IEEE Asia Pacific Conference on Circuits and Systems. (2010).

[15] Samir B. Joshi, J.C.Prajapati, B.H.Soni, "Analysis and Study of FGMOS Based Current Mirror Circuit Using 0.35µm Technology", International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 10, April 2013

[16] W. Jaikla, and M. Siripruchyanun, "Low-component electronically controllable dual-mode universal biquad filter using DO-CCCIIS," Asia-Pacific Conference, pp. 331-334, 2007

[17] Anupama and N. Goel, "Novel DTMOS ±0.5V CCCII with multifunction filter," *International Conference on Computing, Communication & Automation*, Greater Noida, India, 2015, pp. 1232-1237.

[18] E. Farshidi, A. Keramatzadeh, "A New Approach for Low Voltage CMOS based on Current-controlled Conveyors," IJE Transactions B: Applications, vol. 27, no. 5, pp. 723-730, May 2014.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:16:53 UTC from IEEE Xplore. Restrictions apply.

Journal of Fluorescence
https://doi.org/10.1007/s10895-023-03389-8

RESEARCH

Luminescence Studies on Dy^{3+} Doped Calcium Aluminum Borosilicate (CABS) Glasses for White Light Emission and Applications in w-LEDs

Ravita[1] · Aman Prasad[2] · Pooja Rohilla[3] · Rajat Bajaj[3] · Anu[3] · Rajesh Punia[4] · Allam Srinivasa Rao[3]

Received: 6 June 2023 / Accepted: 9 August 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Dy^{3+} doped calcium aluminum borosilicate (CABS) glasses have been synthesized via quick melt quench technique. CABS: xDy^{3+} glasses (x = 0.1, 0.5, 1, 1.5 and 2 mol%) were subjected to various morphological and photoluminescence studies. X-ray diffraction (XRD) and Fourier transform infrared (FT-IR) spectroscopy were conducted to study the structural and bonding nature of the undoped glass. The excitation spectra of Dy^{3+} doped CABS glasses under 574 nm emission show many sharp peaks amongst which the transition from $^6H_{15/2} \rightarrow {}^6P_{7/2}$ (351 nm) had the highest intensity. Under 351 nm excitation, glasses exhibit sharp peaks in the blue, yellow and red regions corresponding to the transitions $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^6H_{13/2}$, $^6H_{11/2}$ and $^6H_{9/2}$ respectively. The dipole–dipole nature of the interaction between the Dy^{3+} ions is confirmed via Dexter theory and Inokuti-Hirayama (I-H) model. CIE coordinates estimated from the emission profiles of these glasses under 351 nm excitation fall in the white region. Considering that these glasses exhibit sharp visible emission under UV excitation, have stable yellow to blue (Y/B) ratios and fast decays with intense energy transfers, we propose to utilise these glasses for white light generation and other white light LED (w-LED) and solid-state lighting (SSL) applications.

Keywords Glasses · Dexter theory · I-H Model · Photoluminescence · w-LEDs · SSL

Introduction

Climate change implications have made one thought clear in today's world that we need to develop more energy efficient resources so that our carbon footprint is systematically reduced leading to a better world for the future generations to live in [1, 2]. This has accelerated research in the development of white light emitting diodes (w-LEDs) that are considered to be fourth generation solid state devices having many favourable properties over the conventional incandescent luminescence sources like high luminous efficiency, low power consumption, longer lifetimes, non-toxicity, eco-friendly and lower production cost [3–8].

Rare earth (RE) ion doped glassy materials are being developed and researched for their immense utility in solid state lighting and w-LED applications [9–11]. RE doped glasses find usage in various applications like 3D display devices, optical fibres, temperature dependent sensors, solid state lasers, optical waveguides, Q-switched devices etc. [12–17]. Different shades of white light can be generated by tuning the red, green and blue (RGB) emissions from the luminescent materials. There are three methods to generate white light [3]: The first method involves usage of some electronic circuit that combines the light coming from RGB LEDs. Second method involves fabrication of a yellow phosphor on a blue emitting LED. Third method involves combining blue and yellow or RGB phosphors with a UV light emitting source. The major disadvantage associated with these methods is the thermal degradation of the epoxy raisin used for encapsulation of phosphors leading to deterioration of luminous efficiency and colour properties [18, 19]. Therefore, it is imperative to develop better alternatives to luminescent phosphors in order to tackle these disadvantages. RE doped glassy materials

✉ Aman Prasad
amanprasad2891@gmail.com

[1] Department of Physics, Chaudhary Bansi Lal University, Bhiwani, Haryana 127021, India

[2] Department of Physics and Computer Science, Dayalbagh Educational Institute (DEI), Deemed University, Dayalbagh, Agra 282005, India

[3] Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

[4] Department of Physics, Maharshi Dayanand University, Rohtak, Haryana 124001, India

Published online: 24 August 2023

are one such alternatives since there is no encapsulation by epoxy raisin involved for fabrication of luminescent material. Moreover, glassy materials are known to have controlled luminescence, high thermal and chemical stability, easy manufacturing process, high solubility for RE ions and exceptional transparency [20, 21].

The choice of host matrix in glass is extremely important as luminescent features are material dependent. Therefore it is very important for the host material to have low phonon energies, low melting point, high thermal and chemical stability and high RE ion solubility [22, 23]. A combination of B_2O_3 and SiO_2 produces borosilicate glasses that has aforementioned characteristics. Borosilicate glasses have properties of both (borate and silicate) glasses have relatively high thermal stability and high transparency than other oxide glasses [9, 10]. Boron forms a strong covalent bond in the glass matrix due to its + 3 oxidation state thereby imparting high thermal stability to the glass. SiO_2 increases the refractive index of the glass thereby increasing the usage of glass in optical applications [24]. But these glasses are marred with high phonon energies that lead to increased non-radiative losses [10, 18, 25]. Alkaline earth metal oxides (CaO) acts as good modifier that will help us to modify the network structure and increases the chemical durability of the borosilicate glasses [26, 27]. Al_2O_3 acts as a network modifier and improves the quality of luminescence by eliminating the possibility of RE ions clustering. It also enhances the thermal and chemical stability of the glass matrix [28, 29].

The trivalent dysprosium ion (Dy^{3+}) is being widely investigated for its ability to generate white light due to intense emissions in the blue and yellow regions. The blue emission corresponds to $^4F_{9/2} \rightarrow {}^6H_{15/2}$ transition which is a pure magnetic dipole interaction and the yellow emission corresponds to the $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transition which is pure electric dipole interaction. The yellow transition is hypersensitive in nature and is affected by the environment around the ion. By tuning the yellow to blue (Y/B) intensity ratio, white light can be generated. The Y/B ratio can be changed by changing the composition of the glass matrix or by changing the ion concentrations [30–32].

In this present research work morphological and luminescent properties of Dy^{3+} doped CABS glasses have been studied in detail. Dexter's theory and Inokuti-Hirayama (I-H) theory curves have been plotted to study the nature of ion-ion interactions. Temperature dependent photoluminescence studies have also been carried out on the as prepared glasses. The encouraging results enforce our idea to use these glasses for white light generation and further w-LED applications.

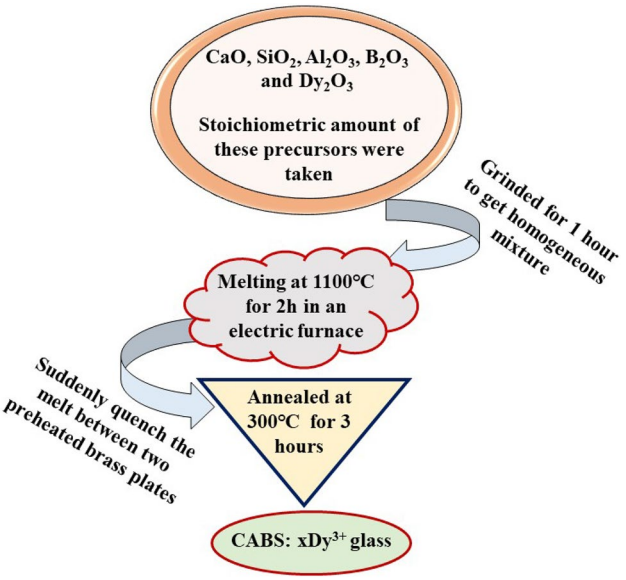


Fig. 1 Schematics of melt-quenching technique

Experimental

Dy^{3+} doped CABS glasses have been synthesized using the conventional melt quenching technique. This process can be understood with the help of schematics shown in Fig. 1. All these precursors used to prepare the aforementioned series were purchased from Sigma-Aldrich with high purity. CABS: xDy^{3+} (x = 0, 0.1, 0.5, 1.0, 1.5 and 2.0) series has been abbreviated as CABS0, CABS1, CABS2, CABS3, CABS4 and CABS5, respectively as shown in Table 1. To check the glassy nature of the titled glasses, we have performed XRD from Bruker D-8. FTIR (Perkin Elmer 2) used to reveal the structural information and presence of several functional groups. The PL studies help us to record the excitation and emission features further helps to understand the energy transfer between Dy-Dy ions. We have done this PL investigation with

Table 1 Composition with sample abbreviations for the as-prepared CABS glasses

Sample ID	B_2O_3	SiO_2	Al_2O_3	CaO	Dy_2O_3
CABS0	50	20	15	15	-
CABS1	50	19.9	15	15	0.1
CABS2	50	19.5	15	15	0.5
CABS3	50	19	15	15	1.0
CABS4	50	18.5	15	15	1.5
CABS5	50	18	15	15	2.0

spectrophotometer (Jasco-FP 8300). The absorption spectrofluorometer (Jasco-V770) used to examine the absorption spectral properties. The PL decay investigation (Edinburgh FL920) has been done to evaluate the lifetime values. The behaviour of PL emission with temperature have been examined form the TDPL instrument from Ocean Optics.

Results and Discussion

Physical and Optical Properties

Table 2 shows the various values of different physical and optical properties of the glass. The density of the as prepared glasses was found out using Archimedes Principle. A Shimadzu electronic balance, having a sensitivity of 10^{-4} mg, was used for the purpose along with double distilled water as an immersion fluid.

The molar volume V_m of the glasses was calculated using the formula [33]:

$$V_m = \frac{Molar\ Mass\ of\ glass\ (M)}{Density\ of\ glass\ (d)} \quad (1)$$

It can be seen from the table that values of density are increasing and molar volume are decreasing with increase in concentration of Dy^{3+} ions in the glass matrix. This can be attributed to replacement of Ca^{2+} ions by the Dy^{3+} ions on account of the larger molecular weights of Dy^{3+} ions.

The refractive indices (n) of the as prepared glasses were found by employing Brewster Angle method. A He–Ne laser of 650 nm wavelength was used for the same. The values are given in the same table. The molar refractivity R_m is defined as total polarizability of 1 mol of a substance and is given by the formula [34]:

$$R_m = \frac{n^2 - 1}{n^2 + 2}(V_m) \quad (2)$$

The electronic polarizability α_m was found out using the value of R_m via the following relation [34]:

$$\alpha_m = \frac{3}{4\pi N} R_m \quad (3)$$

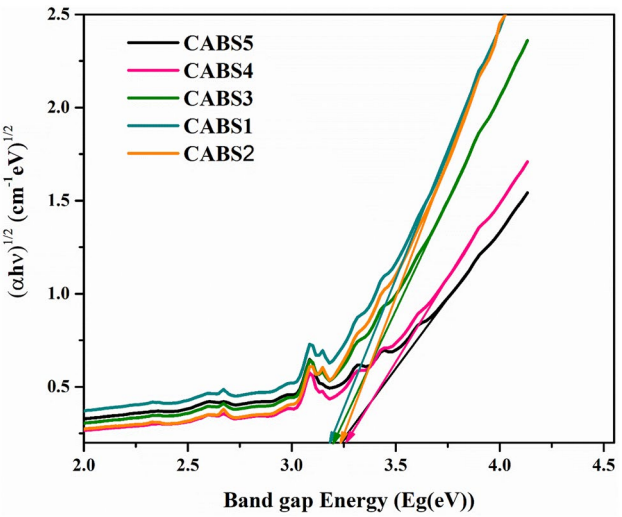


Fig. 2 Tauc plot showing bandgap values for titled glasses

The energy band (E_g) gap values were calculated by drawing a Tauc Plot from the absorption spectra data of the as prepared glasses shown in Fig. 2. The Tauc equation is given as [2]:

$$\alpha h\nu = C(h\nu - Eg)^n \quad (4)$$

where α is the coefficient of absorption and E_g is the energy band gap, hν is the photon energy, C is energy independent constant and n denotes values specific to various transitions i.e., n = 1/2, 2, 1/3 and 3 for direct allowed, indirect allowed, direct forbidden and indirect forbidden transitions respectively. The E_g values range from 3.15 eV to 3.27 eV for the as prepared glasses.

Morphological Properties

The x-ray diffraction (XRD) spectra of the undoped and Dy-doped glasses have been presented in Fig. 3. No sharp peaks are observed in the spectra signifying absence of any long-range order in the glasses. This proves that the host matrix and all Dy-doped glasses are amorphous in nature.

Table 2 Physical properties of Dy^{3+} doped CABS glasses

Sample ID	CABS1	CABS2	CABS3	CABS4	CABS5
Thickness (mm)	2.48	2.48	2.48	2.48	2.48
Average molecular weight	70.74	72.00	73.56	75.12	76.69
Density (g/cm^3)	2.05	2.16	2.33	2.48	2.62
Molar Volume (cm^3)	34.50	33.33	31.57	30.29	29.27
Optical Band Gap (E_g)	3.15	3.23	3.20	3.27	3.23
Refractive Index (n)	2.19	2.16	2.18	2.13	2.17
Molar Refraction R_m (cm^3)	22.59	21.56	20.59	19.34	19.01
Polarizability α_m (10^{-24})	8.95	8.54	8.16	7.66	7.53

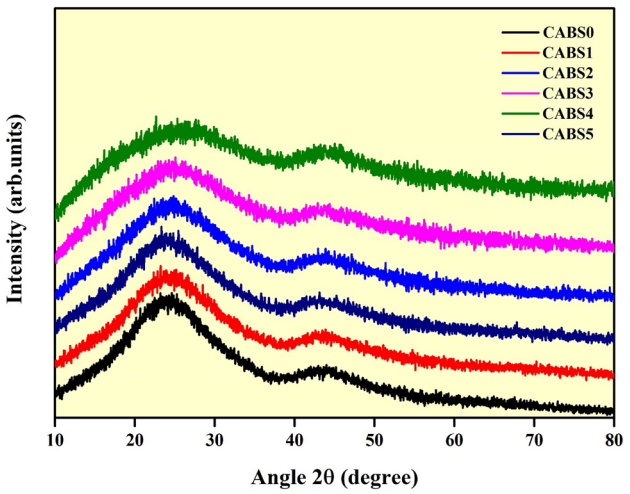


Fig. 3 XRD pattern recorded for CABS glass

The FT-IR spectrum for an undoped CABS glass is shown in Fig. 4. FTIR is a powerful technique to reveal the structural information and functional groups in materials. Here in our case, it contains various shoulders which we have divide into main regions. The first region contains 680, 894, 1073, and 1244 cm^{-1} give information about borate information. The bands observed at 680 cm^{-1} represents the B-O-B linkage and B-O bond [28]. Because of the bending of BO_3-O-BO_4 bond vibrations it gives a band at 894 cm^{-1} [9]. The presence of 1073 cm^{-1} and 1244 cm^{-1} bands corresponds to Si–O-Si linkage and stretching vibrations in BO_4 and BO_3 units [9]. The second regions represent the small peaks due to atmospheric hydrocarbon and carbon dioxide [35]. The less intense band present at 3212 cm^{-1} shows the water content [33].

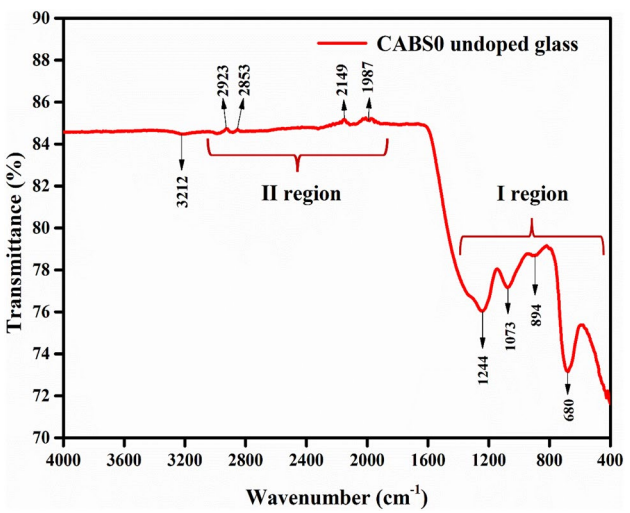


Fig. 4 FT-IR spectrum recorded for an undoped CABS0 glass

Photoluminescence Properties

The excitation spectra of all the CABS glasses are shown in Fig. 5 for 574 nm emission. Eight peaks are visible in the spectra at 325 nm, 339 nm, 351 nm, 365 nm, 388 nm, 427 nm, 453 nm and 474 nm corresponding to the transitions $^6H_{15/2} \rightarrow {}^6P_{3/2}$, $^6H_{15/2} \rightarrow {}^4I_{9/2}$, $^6H_{15/2} \rightarrow {}^6P_{7/2}$, $^6H_{15/2} \rightarrow {}^6P_{5/2}$, $^6H_{15/2} \rightarrow {}^4F_{7/2}$, $^6H_{15/2} \rightarrow {}^4G_{11/2}$, $^6H_{15/2} \rightarrow {}^4I_{15/2}$ and $^6H_{15/2} \rightarrow {}^4F_{9/2}$ respectively [36, 37]. The peak corresponding to $^6H_{15/2} \rightarrow {}^6P_{7/2}$ (351 nm) has the highest intensity so this wavelength was chosen to record the emission spectra of the as prepared glasses.

The emission spectra of all the Dy^{3+} doped CABS glasses is shown in Fig. 6. Four peak at 482 nm (blue), 574 nm (yellow), 662 nm and 703 nm (red) are seen corresponding to the transitions from $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^4F_{9/2} \rightarrow {}^6H_{13/2}$, $^4F_{9/2} \rightarrow {}^6H_{11/2}$ and $^4F_{9/2} \rightarrow {}^6H_{9/2}$ respectively [38, 39]. The peak corresponding to the yellow transition $^4F_{9/2} \rightarrow {}^6H_{13/2}$ is the most intense. It can be seen that the intensity of all the bands increases till 1.0 mol% concentration of Dy^{3+} and then decreases. This can be attributed to the quenching effect whereby increased Dy^{3+}- Dy^{3+} interactions lead to more non-radiative losses in the glass matrix. The inset in Fig. 5. Shows the variation of 575 nm peak intensity with the concentration of Dy^{3+} ions.

Understanding the nature of ion-ion interactions is important in order to study the energy transfer processes taking place in the glass matrix. Ion-ion interaction means the interaction between Dy^{3+} -Dy^{3+} ions. This interaction can be revealed with the help of Dexter's theory [40]. Dexter theory states that luminescence intensity I and molar concentration C, are related as follows [2]:

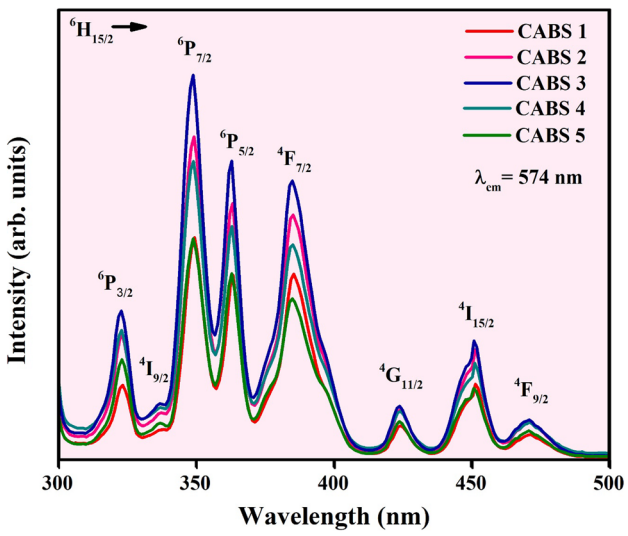


Fig. 5 Excitation spectra recorded for Dy-doped CABS glasses under 574 nm emission wavelength

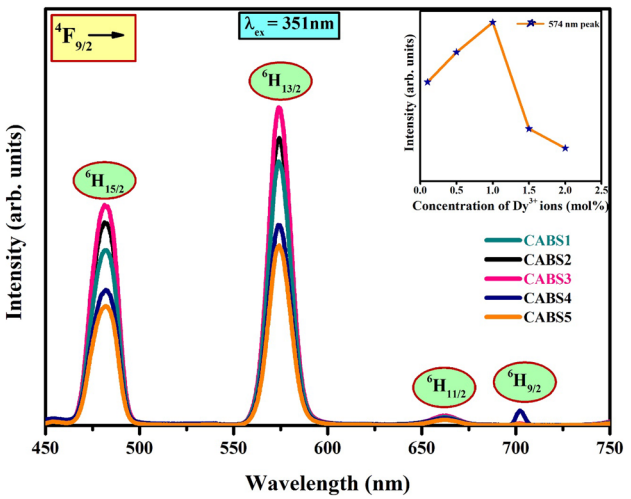


Fig. 6 Emission profiles for CABS glasses recorded under 351 nm excitation wavelength (Inset: Variation of 574 nm peak intensity with the concentration of Dy^{3+} ions)

$$\log\left(\frac{I}{C}\right) = K' - (M/3)\log(C) \tag{5}$$

where K' is a constant and value of M determines whether the ion-ion interactions are dipole–dipole, dipole-quadrupole and quadrupole–quadrupole in nature depending on the value of M being 6,8 or 10 respectively [41]. Figure 7 shows the plot between log(I/C) and log(C). The slope of the graph was used to determine M and its value came out to be 5.82 which is nearly equal to 6. So, it can be deduced that dipole–dipole interactions took place between the Dy^{3+} ions. The intensity of the emission profiles is affected by the value of M. Intensity will be relatively high for M = 6. It means more number of ions are transferring to the acceptor ions. Thus, the nature of ion-ion interactions will lead to somewhere amount of energy transfer found adequate in case of dipole–dipole interaction.

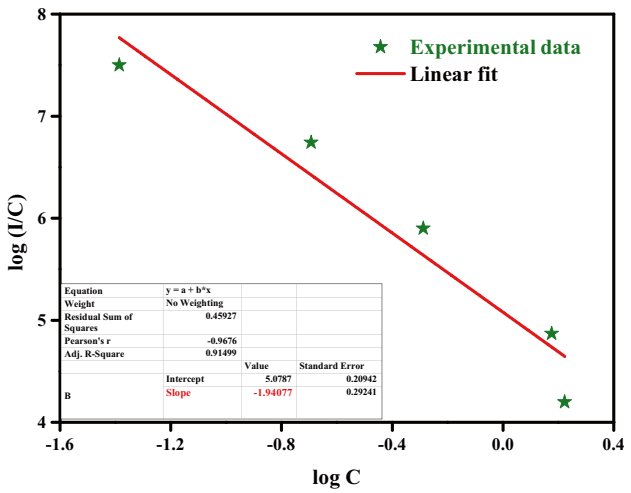


Fig. 7 Dexter's theory shows the energy transfer between Dy^{3+}-Dy^{3+} ions via dipole–dipole interaction in Dy^{3+} doped CABS glasses

A schematic energy level diagram showing various emissions in the glass matrix is given in Fig. 8. It can be seen that all emissions take place from the $^4F_{9/2}$ level. This level gets accumulated with excited Dy^{3+} ions that fall from the upper excited states non-radiatively as the energy difference between these levels is less. Radiative transitions then take place from this level to the lower $^6H_{J/2}$ states (J = 9,11,13,15) due to sufficient energy gap being present between them. The concentration quenching phenomenon can be attributed to the resonance energy transfer (RET) and cross relaxation (CR) channels between the Dy^{3+} ions [40, 42].The possible RET and CR channels can be given as:

$$\text{RET} : (^4F_{9/2} + {}^6H_{15/2} \rightarrow {}^6H_{15/2} + {}^4F_{9/2})$$

$$\text{CR1} : (^4F_{9/2} + {}^6H_{15/2} \rightarrow {}^6H_{9/2} + {}^6F_{3/2})$$

$$\text{CR2} : (^4F_{9/2} + {}^6H_{15/2} \rightarrow {}^6F_{7/2} + {}^6F_{5/2})$$

$$\text{CR3} : (^4F_{9/2} + {}^6H_{15/2} \rightarrow {}^6F_{3/2} + {}^6H_{9/2})$$

The yellow to blue (Y/B) ratio for the as prepared glasses for emission under 351 nm was calculated and presented in Table 3. The Y/B ratio is an important parameter for white light generation. The $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transition is electric dipole ($\Delta J = \pm 2$), which is host sensitive whereas the transition $^4F_{9/2} \rightarrow {}^6H_{15/2}$ is magnetic dipole ($\Delta J = 0, \pm 1$ but $0 \rightarrow 0$ forbidden) and is host insensitive. Thus, the Y/B ratio has been used to determine the covalent nature of the bond between Dy^{3+}-O^{2-}. It can be seen that the ratio lies between 1.41 and 1.46 for changing concentration of Dy^{3+} ions. Higher value of Y/B ratio leads to high covalence. Thus it can be concluded that the local environment around the Dy^{3+} ions is stable and does not get affected with change in the concentration of ions [41–44].

Decay Kinetics

The decay profiles of the as prepared glasses for the 574 nm transition under 351 nm excitation are presented in Fig. 9. The decay curves are found to be bi-exponential in nature and are well fitted to the following expression [45]:

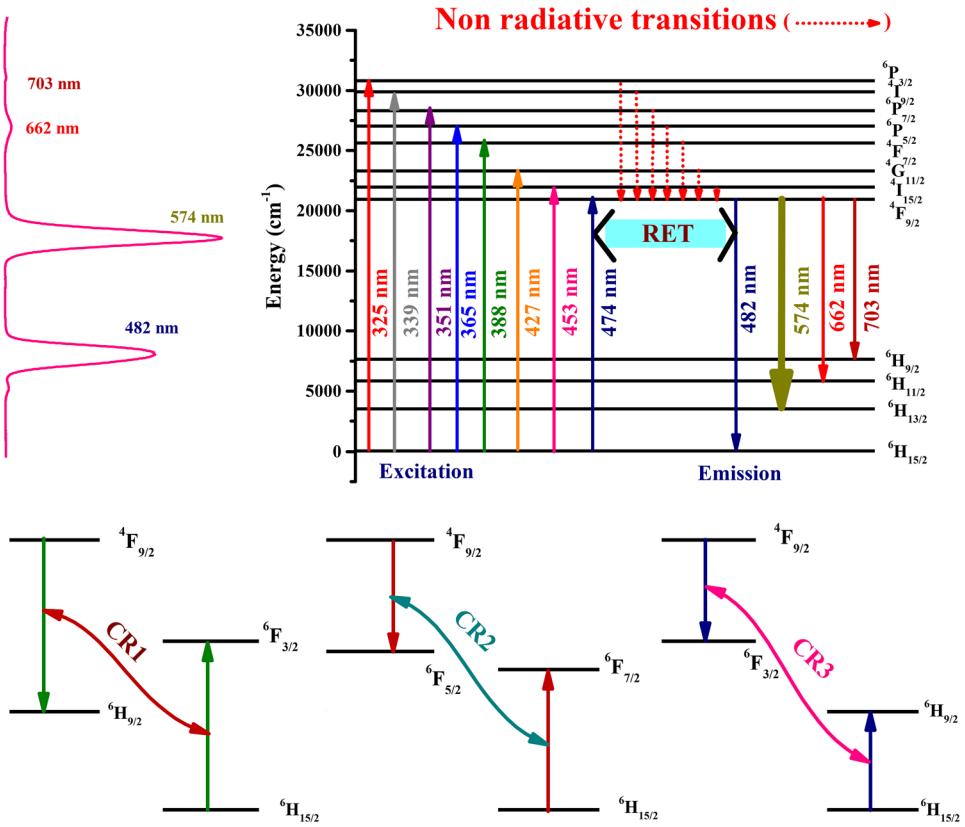


Fig. 8 Partial energy level diagram showing excitation and emission transitions, Cross-relaxation (CR) channels and resonant energy transfer (RET) with non-radiative relaxations for Dy^{3+} doped CABS glasses

Table 3 Average lifetime (τ_{avg}), Energy transfer parameter (Q), acceptor-ion concentration (N_0), critical transfer distance (R_o) and CIE co-ordinates of Dy^{3+} doped CABS glasses

Sample ID	τ_{avg} (ms)	Q	$N_0(10^{19}$ions/cm$^3)$	R_o (Å)	Y/B ratio	CIE coordinates (X, Y)
CABS1	3.93	9.14	0.289	13.27	1.41	0.335, 0.297
CABS2	3.51	14.23	1.56	8.75	1.43	0.349, 0.316
CABS3	3.12	18.68	3.303	7.48	1.46	0.323, 0.317
CABS4	2.84	21.93	5.151	6.80	1.46	0.319, 0.282

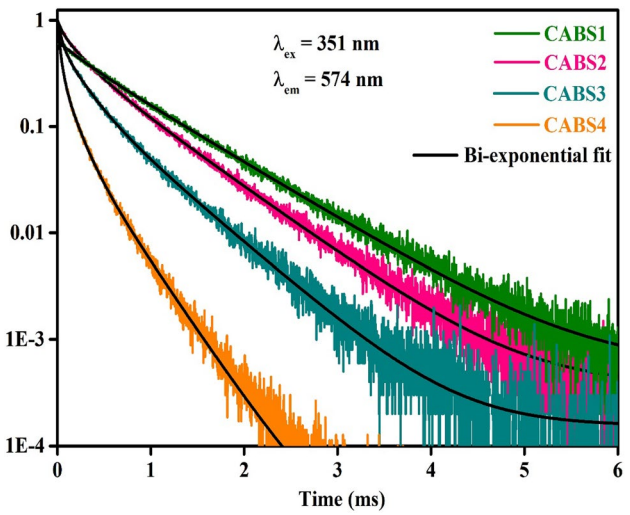


Fig. 9 Decay profiles of Dy^{3+} doped CABS glasses recorded under 351 nm excitation monitoring via $^4F_{9/2} \rightarrow {}^6H_{13/2}$ transition

$$I = I_0 + A\ exp(-t/\tau_1) + B\ exp(-t/\tau_2) \quad (6)$$

where I_0 and I are the luminescence intensities at zero and t time, A and B are the fitting constants, τ_1 and $\tau 2$ are fast and slow components of lifetime. The average lifetime values can be estimated by the following expression [46]:

$$\tau_{av} = \frac{A\tau_1^2 + B\tau_2^2}{A\tau_1 + B\tau_2} \quad (7)$$

These values are shown in Table 3. It can be seen that the values of lifetime decrease with the increase in the concentration of Dy^{3+} ions. This can be due to the increased energy transfers between one Dy^{3+} ion and neighbouring Dy^{3+} ions in the glass matrix with the increase in their concentration.

In order to justify the dipole–dipole nature of interaction between the Dy^{3+} ions in the glass matrix, Inokuti-Hirayama model was applied to the decay curve of CABS3

glass. This model states that the emission intensity is given as [47]:

$$I(t) = Ioexp\left[-\frac{t}{\tau o} - Q(\frac{t}{\tau o})^{3/S}\right] \quad (8)$$

where t is the time after excitation and t_0 is the intrinsic decay time of the donors in the absence of acceptors. S can take either 6, 8 or 10 depending upon whether the ion-ion interactions are dipole–dipole or dipole-quadrupole or quadrupole–quadrupole respectively. Q is the energy transfer parameter given by the following relation:

$$Q = \frac{4\pi}{3}\Gamma\left(1 - \frac{3}{S}\right)NoRo^3 \quad (9)$$

Here Γ(x) is 1.77 for dipole–dipole interactions, for dipole-quadrupole interactions it is 1.43 and for quadrupole–quadrupole interactions, it is 1.3. N_0 is the concentration of rare earth ions and R_0 is defined as the critical energy transfer distance between the donor and acceptor ions.

It can be seen from Fig. 10 that the curves fit well for S = 6 signifying the dipole–dipole interaction between the Dy^{3+} ions in the host matrix. The values of Q were estimated from I-H profiles and used to calculate the value of R_0 from Eq. (9). All these values are presented in Table 3. It is evident that the values of Q increase with increasing concentration of Dy^{3+} thereby signalling more energy transfer between the Dy^{3+} ions and hence a faster decay time. The value of R_0 decreases with the increase in N_0 also supports the increment in Q parameter. As the distance between donor and acceptor decreases, the possibility of energy transfer increases [9]. More energy transfer leads to the higher emission intensity.

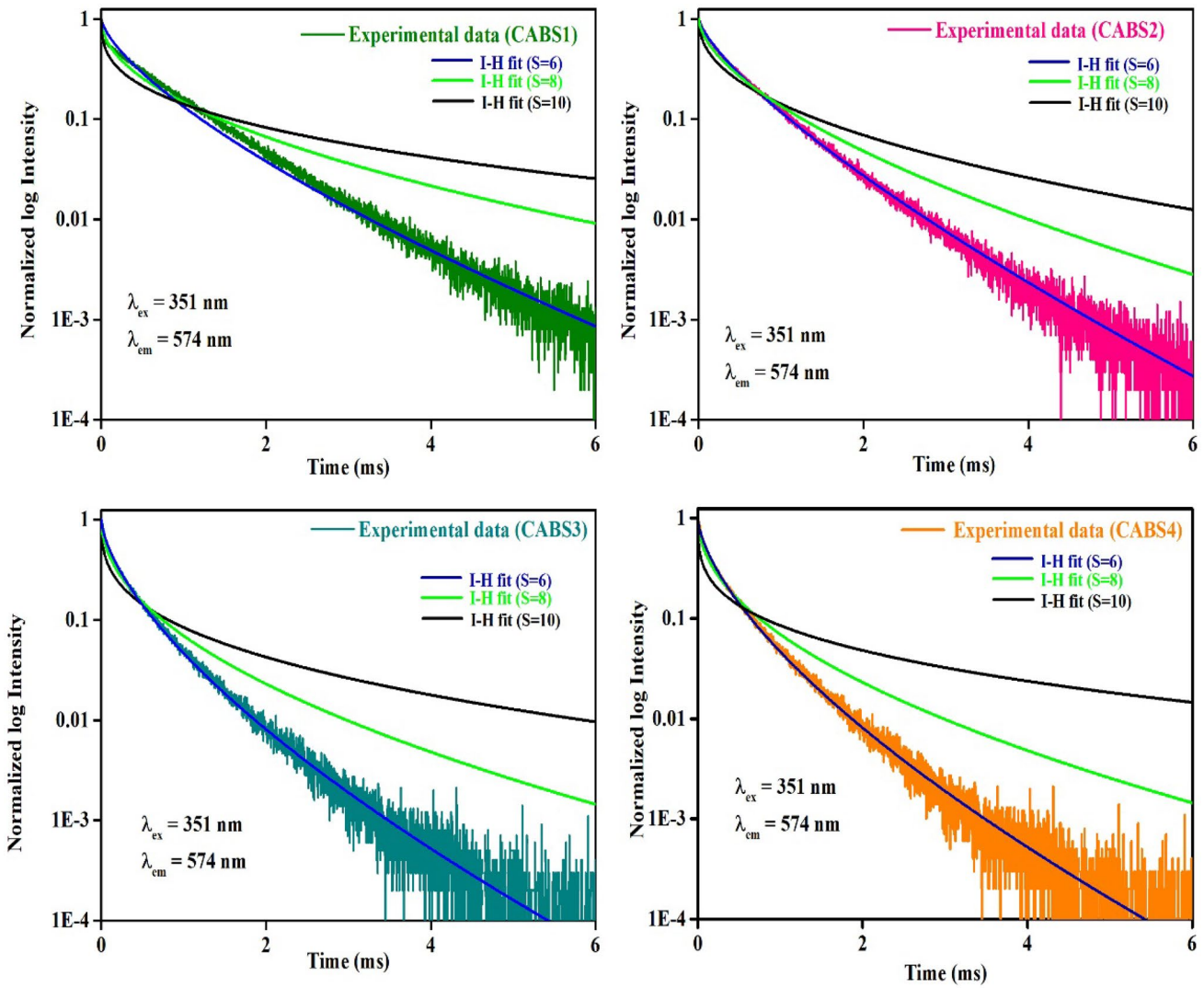


Fig. 10 Decay profiles fitted with I-H model for Dy doped CABS glasses under 351 nm excitation wavelength

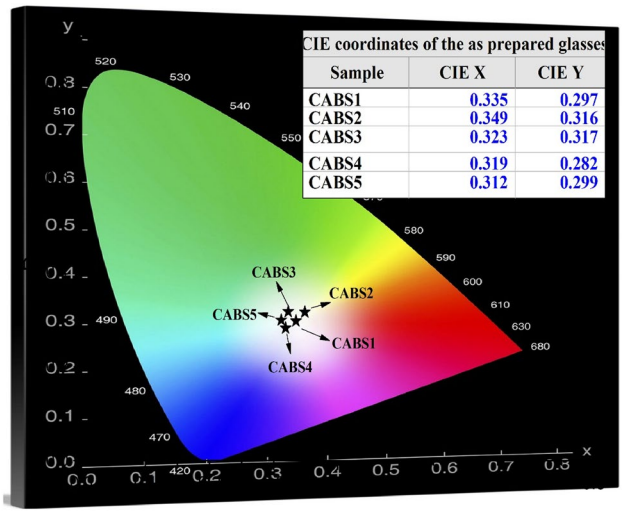


Fig. 11 CIE chromaticity co-ordinates for Dy^{3+} doped CABS glasses under 351 nm excitation wavelength

Colorimetry Analysis

The CIE chromaticity coordinates for the emission spectra under 351 nm excitation have been calculated and presented in Table 3. The same coordinates have been shown on a CIE diagram in Fig. 11. As is evident from the figure, all the coordinates lie in the white region of the diagram and are pretty close to the standard white light point (0.333, 0.333). Thus, Dy^{3+} doped CABS glasses can indeed be used for white light generation and other SSL applications.

Conclusions

Dy^{3+} doped calcium alumino borosilicate glasses (CABS) have been prepared via quick melt quenching technique. Morphological and photoluminescence studies have been carried out extensively in the as prepared glasses. Under 351 nm excitation, these glasses emit sharp bands in the blue and yellow regions along with feeble bands in the red regions corresponding to the transitions $^4F_{9/2} \rightarrow {}^6H_{15/2}$, $^6H_{13/2}$, $^6H_{11/2}$ and $^6H_{9/2}$ respectively. Dexter theory and Inokuti-Hirayama model proves that the Dy^{3+}- Dy^{3+} ion interaction is dipole–dipole in nature. The decay profiles show fast decay with increasing concentration of Dy^{3+} ions. The CIE coordinates of these glasses for the emission spectra under 351 nm excitation fall well within the white region. These glasses can be suitably utilised for white light emission and further usage in w-LEDs and SSL applications.

Author Contributions All authors contributed to the study, conception and design of the manuscript. The author credits are as follows: 1. Ravita: Conceptualisation, synthesis, data analysis, writing. 2. Aman Prasad: Data analysis, review. 3. Pooja Rohilla: Characterisation, data analysis. 4. Rajat Bajaj: Synthesis, characterisation. 5. Anu: Data analysis. 6. Rajesh Punia: Review. 7. Allam Srinivasa Rao: Conceptualisation, supervision, analysis, review.

Data Availability The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

Declarations

Animal Ethics The authors declare that no animal or animal data was used in any study or experiment mentioned in this manuscript.

Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Kaur S, Katyal V, Plakkot V et al (2021) Radiative emission analysis of Sm^{3+} ions doped borosilicate glasses for visible orange photonic devices. J Non Cryst Solids 572:121106. https://doi.org/10.1016/j.jnoncrysol.2021.121106
2. Bajaj R, Prasad A, Yeswanth AVS et al (2021) Down-shifting photoluminescence studies of thermally stable Dy^{3+} ions doped borosilicate glasses for optoelectronic device applications. J Mater Sci Mater Electron. https://doi.org/10.1007/s10854-021-07667-8
3. Jose A, Krishnapriya T, George A et al (2021) Cool/warm white light luminescent traits and energy transfer studies of Dy^{3+} / Er^{3+} / Sm^{3+} triply doped multicomponent borosilicate glasses for lighting applications. J Non Cryst Solids 562:120775. https://doi.org/10.1016/j.jnoncrysol.2021.120775
4. Hao Y, Cao J (2016) Structure and luminescence of Dy^{3+} doped $CaO–B_2O_3–SiO_2$ glasses. Phys B Condens 493:68–71. https://doi.org/10.1016/j.physb.2016.04.017
5. Farooq U, Zhao Z, Sui Z et al (2019) Tm^{3+}/Dy^{3+}/Eu^{3+} (Sm^{3+}) tri-activated Y_2WO_6 as one potential single-phase phosphor for WLEDs. J Alloys Compd 778:942–950. https://doi.org/10.1016/j.jallcom.2018.11.091
6. Ren Q, Zhao Y, Wu X et al (2021) Luminescence characteristics of a novel Tm/Eu co-doped polychromatic tunable phosphor. J Solid State Chem 294:121869. https://doi.org/10.1016/j.jssc.2020.121869
7. Bashar KA, Fong WL, Haider KA et al (2020) Optical studies on Tb^{3+} : Dy^{3+} singly and doubly doped Borosilicate glasses for white light and solid state lighting applications. J Non Cryst Solids 534:119943. https://doi.org/10.1016/j.jnoncrysol.2020.119943
8. Yu M, Xu X, Zhang W et al (2020) The effect of Sm^{3+} co-doping on the luminescence properties of $Ca_{2.85}Li_{0.15}(PO_4)_{1.85}(SO_4)_{0.15}$: Dy^{3+} white-emitting phosphors. J Alloys Compd 817:1–9. https://doi.org/10.1016/j.jallcom.2019.152761
9. Ravita, Rao AS (2021) Effective sensitization of Eu^{3+} visible red emission by Sm^{3+} in thermally stable potassium zinc alumino borosilicate glasses for photonic device applications. J Lumin 118689. https://doi.org/10.1016/j.jlumin.2021.118689
10. Deopa N, Kumar B, Sahu MK et al (2019) Effect of Sm^{3+} ions concentration on borosilicate glasses for reddish orange

luminescent device applications. J Non Cryst Solids 513:152–158. https://doi.org/10.1016/j.jnoncrysol.2019.03.025
11. Jose A, Krishnapriya T, Jose TA et al (2021) Scripta Materialia Color tunable luminescence characteristics and energy transfer analysis of Dy^{3+} / Sm^{3+} co-doped multicomponent borosilicate glasses. Scr Mater 203:114088. https://doi.org/10.1016/j.scriptamat.2021.114088
12. Lopez-Herrer HU, Escorcia-Garcia J, Gutierrez-Chavarria CA et al (2022) Influence of CaO and Dy2O3 on the structural, chemical and optical properties of thermally stable luminescent silicate nanoglass-ceramics. J Mater Res Technol 20:1544–1557. https://doi.org/10.1016/j.jmrt.2022.07.148
13. Hegde V, Devarajulu G, Pramod AG et al (2022) Analysis of optical and near-infrared luminescence of Er^{3+} and Er^{3+} / Yb^{3+} co-doped heavy metal borate glasses for optical amplifier applications. Photonics 9:355. https://doi.org/10.3390/photonics9050355
14. Ramesh P, Gangareddy J, Sathish KN et al (2021) Effect of heavy metal oxides on photoluminescence and spectroscopic attributes of Eu^{3+} activated borate glasses. Opt Mater (Amst) 114:110933. https://doi.org/10.1016/j.optmat.2021.110933
15. Bhemarajam J, Babu MM, Prasad PS, Prasad M (2021) Spectroscopic studies on Er^{3+} ions incorporated bismuth borolead lithium glasses for solid state lasers and fiber amplifiers Spectroscopic studies on Er^{3+} ions incorporated bismuth borolead lithium glasses for solid state lasers and fiber amplifiers. Opt Mater. https://doi.org/10.1016/j.optmat.2021.110818
16. Wantana N, Kaewnum E, Damdee B et al (2017) Energy transfer based emission analysis of Eu^{3+} doped Gd_2O_3-CaO-SiO_2-B_2O_3 glasses for laser and X-rays detection material applications. J Lumin 194:75–81. https://doi.org/10.1016/j.jlumin.2017.10.004
17. Wan X, Lin Y, Tie S, Shen J (2011) Luminescence and energy transfer in Dy^{3+}/Tb^{3+} co-doped CaO–Al_2O_3–B_2O_3–RE_2O_3 glass. J Non Cryst Solids 357:3424–3429. https://doi.org/10.1016/j.jnoncrysol.2011.06.011
18. Ravina Naveen, Sheetal, et al (2021) Judd-Ofelt itemization and influence of energy transfer on Sm^{3+} ions activated B_2O_3–ZnF_2–SrO–SiO_2 glasses for orange-red emitting devices. J Lumin 229:117651. https://doi.org/10.1016/j.jlumin.2020.117651
19. George H, Deopa N, Kaur S et al (2019) Judd-Ofelt parametrization and radiative analysis of Dy^{3+} ions doped Sodium Bismuth Strontium Phosphate glasses. J Lumin 215:116693. https://doi.org/10.1016/j.jlumin.2019.116693
20. Qiao Y, Chen P (2021) Luminescence, energy transfer, and color adjustment of CaO-CaF_2 -Al_2O_3 -B_2O_3-SiO_2 glass co-doped with CeO_2 and Sm_2O_3. J Non Cryst Solids 552:120461. https://doi.org/10.1016/j.jnoncrysol.2020.120461
21. Kumar M, Rao AS (2020) Concentration-dependent reddish-orange photoluminescence studies of Sm^{3+} ions in borosilicate glasses. Opt Mater (Amst) 109:110356. https://doi.org/10.1016/j.optmat.2020.110356
22. Peddaiah M, Salma C, Rudramadevi BH (2021) Optik Structural and an orange-red emission studies of Sm^{3+} doped $Ba_3La_2(BO_3)_4$ phosphor for solid state lighting application. Optik (Stuttg) 244:166695. https://doi.org/10.1016/j.ijleo.2021.166695
23. Kumar M, Rao AS (2021) Influence of Tb^{3+} ions concentration and temperature on lithium bismuth alumino borosilicate glasses for green photonic device applications. Opt Mater (Amst) 120:111439. https://doi.org/10.1016/J.OPTMAT.2021.111439
24. Zhang X, Xu T, Nie Q et al (2007) Spectroscopic properties of Er^{3+} / Yb^{3+} co-doped Bi_2O_3 B_2O_3 GeO_2 glasses. Spectrochim Acta 67:246–250
25. Sailaja P, Mahamuda S, Swapna K et al (2021) Broadband NIR emission at 1.53 µm in trivalent erbium ions doped SrO-Al_2O_3-B_2O_3-$BaCl_2$–10TeO_2 glasses for optical fiber and NIR laser applications. J Non Cryst Solids 567:120937. https://doi.org/10.1016/j.jnoncrysol.2021.120937
26. Abdel-Aziz AM, Elsad RA, Ahmed EM et al (2021) Physical, FTIR, ultrasonic, and dielectric characteristics of calcium lead-borate glasses mixed by Nd_2O_3/Er_2O_3 rare earths: experimental study. J Mater Sci Mater Electron 32:19966–19979. https://doi.org/10.1007/s10854-021-06521-1
27. Li C, Zhang X, Onah VC et al (2022) Influence of alkali metal and alkaline earth metal oxides on physical and optical properties of TeO_2–WO_3–GdF_3-R_2O-ZO: Eu^{3+} glasses (R = Li, Na, K, Z = Mg, Ca, Ba). Ceram Int 48:28718–28729. https://doi.org/10.1016/j.ceramint.2022.06.186
28. Ahmed MR, Ashok B, Ahmmad SK et al (2019) Infrared and Raman spectroscopic studies of Mn^{2+} ions doped in strontium alumino borate glasses: Describes the role of Al_2O_3. Spectrochim Acta - Part A Mol Biomol Spectrosc 210:308–314. https://doi.org/10.1016/j.saa.2018.11.053
29. Kumar A, Anu SMK, Ravita DS, Deopa N, Malik Z, Punia R, Rao AS (2022) Spectral characteristics of Tb^{3+} doped ZnF_2–K_2O–Al_2O_3–B_2O_3 glasses for epoxy free tricolor w-LEDs and visible green laser applications. J Lumin 244:118676
30. Chanthima N, Tariwong Y, Kaewkhao J et al (2021) White Light Emission of Bi_2O_3 -B_2O_3 -P_2O_5 Glasses Doped with Dy^{3+} Ion. Mater Today Proc 43:2498–2507. https://doi.org/10.1016/j.matpr.2020.04.609
31. Liu R, Chen M, Zhu X et al (2020) Luminescent properties and structure of Dy^{3+} doped germanosilicate glass. J Lumin 226:117378. https://doi.org/10.1016/j.jlumin.2020.117378
32. Mahamuda S, Swapna K, Packiyaraj P et al (2014) Lasing potentialities and white light generation capabilities of Dy^{3+} doped oxy- fluoroborate glasses. J Lumin 153:382–392. https://doi.org/10.1016/j.jlumin.2014.03.009
33. Ali AA, Rammah YS, El-Mallawany R, Souri D (2017) FTIR and UV spectra of pentaternary borate glasses. Meas J Int Meas Confed 105:72–77. https://doi.org/10.1016/j.measurement.2017.04.010
34. El-Damrawi G, Abdelghany AM, Oraby AH (2020) Structural and optical absorption studies on Cr_2O_3 doped SrO-P_2O_5 glasses. Spectrochim Acta - Part A Mol Biomol Spectrosc 228:117840. https://doi.org/10.1016/j.saa.2019.117840
35. Shasmal N, Molla AR, Karmakar B (2014) Synthesis and characterization of chloroborosilicate glasses in the K_2O-BaO-Al_2O_3-B_2O_3-SiO_2-$BaCl_2$ system. J Non Cryst Solids 398–399:32–41. https://doi.org/10.1016/j.jnoncrysol.2014.04.017
36. Reddi Babu M, Madhusudhana Rao N, Mohan Babu A et al (2016) Effect of Dy^{3+} ions concentration on optical properties of lead borosilicate glasses for white light emission. Optik (Stuttg) 127:3121–3126. https://doi.org/10.1016/j.ijleo.2015.12.018
37. Poonam Shivani, Anu, et al (2020) Judd-Ofelt Parameterization and Luminescence Characterization of Dy^{3+} Doped Oxyfluoride Lithium Zinc Borosilicate Glasses for Lasers and w-LEDs. J Non Cryst Solids 544:120187. https://doi.org/10.1016/j.jnoncrysol.2020.120187
38. Kiran N, Suresh Kumar A (2013) White light emission from Dy^{3+} doped sodium-lead borophosphate glasses under UV light excitation. J Mol Struct 1054–1055:6–11. https://doi.org/10.1016/j.molstruc.2013.09.023
39. Babu AM, Jamalaiah BC, Kumar JS et al (2011) Spectroscopic and photoluminescence properties of Dy^{3+}-doped lead tungsten tellurite glasses for laser materials. J Alloys Compd 509:457–462. https://doi.org/10.1016/j.jallcom.2010.09.058
40. Sahu MK, Mula J (2019) White light emitting thermally stable bismuth phosphate phosphor $Ca_3Bi(PO_4)_3$:Dy^{3+} for solid-state lighting applications. J Am Ceram Soc 102:6087–6099. https://doi.org/10.1111/jace.16479
41. Kaur S, Vishwakarma AK, Deopa N et al (2018) Spectroscopic studies of Dy^{3+} doped borate glasses for cool white light generation. Mater Res Bull 104:77–82. https://doi.org/10.1016/j.materresbull.2018.04.002

42. Jha K, Jayasimhadri M (2016) Spectroscopic investigation on thermally stable Dy^{3+} doped zinc phosphate glass for white light emitting diodes. J Alloys Compd 688:833–840. https://doi.org/10.1016/j.jallcom.2016.07.024
43. Vijayakumar R, Guo H, Huang X (2018) Energy transfer and color-tunable luminescence properties of Dy^{3+} and Eu^{3+} co-doped $Na_3Sc_2(PO_4)_3$ phosphors for near-UV LED-based warm white LEDs. Dye Pigment 156:8–16. https://doi.org/10.1016/j.dyepig.2018.03.053
44. Jaidass N, Krishna Moorthi C, Mohan Babu A, Reddi Babu M (2018) Luminescence properties of Dy^{3+} doped lithium zinc borosilicate glasses for photonic applications. Heliyon 4:1–15. https://doi.org/10.1016/j.heliyon.2018.e00555
45. Rani PR, Venkateswarlu M, Mahamuda S et al (2018) Structural, absorption and photoluminescence studies of Sm^{3+} ions doped barium lead alumino fluoro borate glasses for optoelectronic device applications. Mater Res Bull 110:159–168. https://doi.org/10.1016/j.materresbull.2018.10.033
46. Li K, Liu X, Zhang Y et al (2015) Host-sensitized luminescence properties in $CaNb_2O_6:Ln^{3+}$ (Ln^{3+} = Eu^{3+}/Tb^{3+}/Dy^{3+}/Sm^{3+}) phosphors with abundant colors. Inorg Chem 54:323–333. https://doi.org/10.1021/ic502493c
47. Prasad A, Rao AS, Prakash GV (2020) A study on up-conversion and energy transfer kinetics of $KGdF_4$: Yb^{3+} / Er^{3+} nanophosphors. J Mol Struct 1205:127647. https://doi.org/10.1016/j.molstruc.2019.127647

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Int. Journal of Business Science and Applied Management, Volume 18, Issue 2, 2023

Mapping the Field of Research on Entrepreneurial Success: A Bibliometric Study and Future Research Agenda

Varda Sardana
Jaipuria Institute of Management
A-32A, Opposite IBM India, Sector 62, Noida, 201309, India
Tel: 9654435464
Email: varda.sardana@jaipuria.ac.in

Shubham Singhania
University School of Management and Entrepreneurship, Delhi Technological University
M893+JHM, Vivek Vihar Phase 2, Block E, Vivek Vihar, Delhi, 110095, India
Tel: 9711924415
Email: shubhamsinghania1@gmail.com

Abstract
Entrepreneurship has lately become a buzzword, and emphasis is being placed on ensuring the success of entrepreneurial ventures through both financial and non-financial means. This study aims to critically review and provide insights into the key developments in the literature on entrepreneurial success. The review employs bibliometric methods and content analysis on 595 research articles extracted from the Scopus Database for the period 1996 to 2021. The findings of the study highlight the growth trends of publications on entrepreneurial success, the most impactful articles, journals, and countries working in this domain as well as the evolution of entrepreneurial research topics over the years. Six clusters, representing major themes that have been explored in the literature in the past 26 years, are uncovered through content analysis. Future research agendas have been elaborated based on the emerging topics of study.

Keywords: entrepreneurship, entrepreneurial success, bibliometric analysis, literature review

1. INTRODUCTION

Entrepreneurship refers to the identification of opportunities in order to create new business ventures, provide goods and services (Shane and Venkataraman, 2000), generate employment, and promote innovation. It acts as a major force in the economic growth of a country (Van Praag and Versloot, 2007). Entrepreneurial development not only leads to employment generation but also leads to social development and improvement in the global competitiveness of a country (Naudé, 2013). Entrepreneurial success may be defined as the ability of the firm to be able to survive in the market over the long term (Fisher et al., 2014). Such success can be viewed from two angles: financial success and non-financial success (Orser and Dyke, 2009; Gorgievski et al., 2010). Financial success can be measured with the help of organizational economic parameters such as firm survival rate, cash flows, company sales and profits, the number of employees, and the company growth rate (Dej, 2010). Non-financial success, on the other hand, is based on the personal attributes of the entrepreneurs, such as personal fulfillment, societal contributions, self-realizations, and work-related social relationships (Dej, 2010). Financial success is often argued to represent the short-term outlook, whereas non-financial or psychological success is attributed to the long-term perspective (Simon et al., 2015). Literature suggests that the success of a startup may hold different meanings for different entrepreneurs. While for some, it may boil down to the achievement of the enterprise's goals and objectives (Hyder and Lussier, 2016) or a continuous high financial performance (Spiegel et al., 2015), for others, a startup may be successful on being acquired by another company at a good valuation (Krejci et al., 2015).

The process of entrepreneurship involves three phases: opportunity identification phase, development and execution phase, and survival and growth phase (Baron and Shane, 2007). The success of an entrepreneurial venture is mostly evaluated during the second phase. Various indicators are taken into consideration while evaluating a venture's success. One of the most paramount indicators is financial gain (Parker, 2009), while other indicators include sales, revenue, and market development (Rauch and Frese, 2007). It is difficult to define entrepreneurial success based on financial gains solely because many start-ups may not be able to become profitable initially, even when the firm enjoys increasing sales, due to high set-up costs and interest payments (Perez and Canino, 2009). Still, it is an integral indicator as low financial gains or profitability can prove to be dangerous for a business venture and might lead to its death (Coad, 2007). Contrary to the above, a firm may dissolve even if it is profitable but is unable to fulfill personal goals (Green et al., 2003).

Various factors drive entrepreneurial success, which can be classified into external and internal drivers. The external factors include economic factors and social factors, while the internal factors comprise psychological factors and personal factors. Economic factors comprise access to financial resources (Gupta and Mirchandani, 2018), an increase in sales and revenues, continuous development, customer acceptance and satisfaction, and the overall economic growth of the firm (Wach et al., 2016). The social factors include 2 major components: workplace relationships, that is, the ability to maintain healthy relationships with the stakeholder within and outside the organization (Jayawarna et al., 2013), and societal impact, which refers to the impact of the firm's products and services on the environment as well as on the health and well-being of the society (Florea et al., 2013). Psychological factors comprise components such as resilience, self-efficacy, and the optimism of the entrepreneurs that influences their behaviors, intentions, and motivations, (Baluku et al., 2016). Resilience is the ability of the entrepreneur to bounce back from unprecedented, adverse circumstances by adapting to a dynamic environment and bringing positive change (Luthans et al., 2007), while self-efficacy is the confidence of an entrepreneur to develop and build a new venture, accomplish the goals, overcome all the challenges, and be proficient in the entrepreneurial journey (McGee et al., 2009; Newman et al., 2014). Optimism is a psychological factor that motivates an entrepreneur by increasing the possibility of positive outcomes (Luthans et al., 2007; Trevelyan, 2008). Lastly, the set of personal factors comprises the education level of the entrepreneurs (Barreneche, 2014; Kolstud and Wiig, 2015), their innovativeness and creativity (Mazzucato, 2011), and their personal initiative as well as active actions towards ensuring the survival and growth of the venture (Frese and Gielnik, 2014).

While in the past, emphasis has been on economic as well as psychological parameters for determining the growth and development of businesses, recent literature has focused attention on the emotional skills of the entrepreneurs (Aly et al., 2021), which are internal in nature, as well as family-related factors (Staniewski and Awruk, 2021), which is an external factor. Dimensions such as parental attitudes, the manner of communication amongst the family, and the origin of the family have now been recognized as playing a crucial role in the entrepreneurial success (Staniewski and Awruk, 2021).

Despite the growing number of studies on entrepreneurial start-ups, the literature suffers from major gaps in quantitatively identifying the intellectual developments and evolution taking place in the domain of entrepreneurial success. The topic of entrepreneurial success is emerging but is under-researched (Fisher et al., 2014). Though there are fragmented studies that focus on various aspects of entrepreneurial success, such as achievement motivation (Staniewski and Awruk, 2019), age (Zhao et al., 2021), the relevance of resilience (Salisu et al., 2021), the literature still lacks a quantitative review to examine the growth in this domain.

The present study is an attempt to holistically fill this gap and provide valuable insights into the key developments that have taken place in this domain and underline future themes for research. When the volume of publications becomes sufficiently large over a period, it is not possible to review the literature qualitatively alone and, in such cases, bibliometric analysis becomes the imperative solution. To bring continuity into the entrepreneurial research domain and to give future directions, we address the following research questions:

RQ1: What are the spatial and geographical trends as well as the volume of publications on entrepreneurial success?

RQ2: Which are the top journals publishing in the domain of the success of entrepreneurial ventures?

RQ3: Which are the most influential research articles in this research area?

RQ4: Who are the top authors publishing in the domain of entrepreneurial success?

RQ5: Which topics in the entrepreneurial success literature have been trending in the past and which topics are emerging?

RQ6: What are the key themes that have been explored by researchers in the domain of entrepreneurial success and what are the future avenues for research?

The above questions are answered by analyzing 595 peer-reviewed research articles and review papers extracted from the Scopus Database. The review employs bibliometric methods (Boyack and Klavans, 2010) and content analysis to discover the evolution and intellectual development in the domain of entrepreneurial success. Previous studies in various domains such as corporate governance (Singh et al., 2021a), risk management (Singhania et al., 2022), and deposit insurance (Sardana and Singhania, 2022) have also made use of bibliometric analysis to answer such types of research questions.

2. LITERATURE REVIEW

Many small businesses and start-ups evolve every year, but more than a quarter of them dissolve within the first year of their operation (McKenzie and Paffhausen, 2018). Out of the start-ups that survive, a small proportion can grow and contribute to the generation of employment and the development of the economy (Olafsen and Cook, 2016). The success rate of the start-ups is, therefore, quite low, and it is difficult to predict which of the evolving firms will be more successful (Nanda, 2016). Entrepreneurship is a risky yet exciting journey that requires not just ideas and funds, but ability, skills, determination, persistence, hard work, and most importantly, patience. This package deal leads to the path toward entrepreneurial success (Woodfield et al., 2017).

Traditionally, entrepreneurial success has been considered solely based on financial and economic indicators (Zhou et al., 2017) such as profits, probability of survival, and size of the firm (Fried and Tauer, 2015). But the importance of non-financial indicators cannot be ignored (Razmus and Laguna, 2018). As far as the factors driving entrepreneurial success are concerned, the skills and knowledge possessed by an entrepreneur (Abu et al., 2014), the entrepreneur's psychological behaviour, proficiency, ability to manage work stress, and commitment towards a task (Gupta and Mirchandani, 2018), the internal locus of control of an entrepreneur (Asante and Affum-Osei, 2019), entrepreneurs' attitude, behaviour and mentality (Saptono and Najah, 2018) have gained importance in theory as well as practice. In order to ensure the smooth functioning of the organization and to deal with uncertain circumstances, the cultural attitudes of entrepreneurs (Yusof et al., 2017), appreciation and strong implementation of work-life balance (Orlandi, 2017), autonomy for communication, accepting behaviour towards challenges, proactiveness and a decision-making attitude (Toms et al., 2019), innovativeness (Ojo et al., 2017), and the availability of a well-written and researched business plans (Agarwal and Dahm, 2015) are also essential. Another important factor is gender (Muis et al., 2017) as different genders possess different characteristics, and the success of the venture has varying meanings for different genders.

2.1 Theories of Entrepreneurial Success

The concept of entrepreneurial success is built on theories taken from different disciplines. Some of the major theories include cognitive theory, discovery theory, creation theory, Dubin's theory, and the theory of economic development. The cognitive theory takes into account all those internal processes that help an individual identify the knowledge structures around them to make decisions, assessments, judgments, etc. (Cacciolatti, et al., 2015). This theory acts as a guiding force for entrepreneurs and determines their risk-taking behaviour, which differs significantly from non-entrepreneurs (Palich et al., 1995). Discovery theory is based on the premise that the markets are dynamic, and this gives room for entrepreneurs to discover the existing opportunities in the markets as well as tap them for the generation of profits (Kessler, 2013). On the other hand, creation theory is of the view that opportunities are not readily available in the market, rather they are the

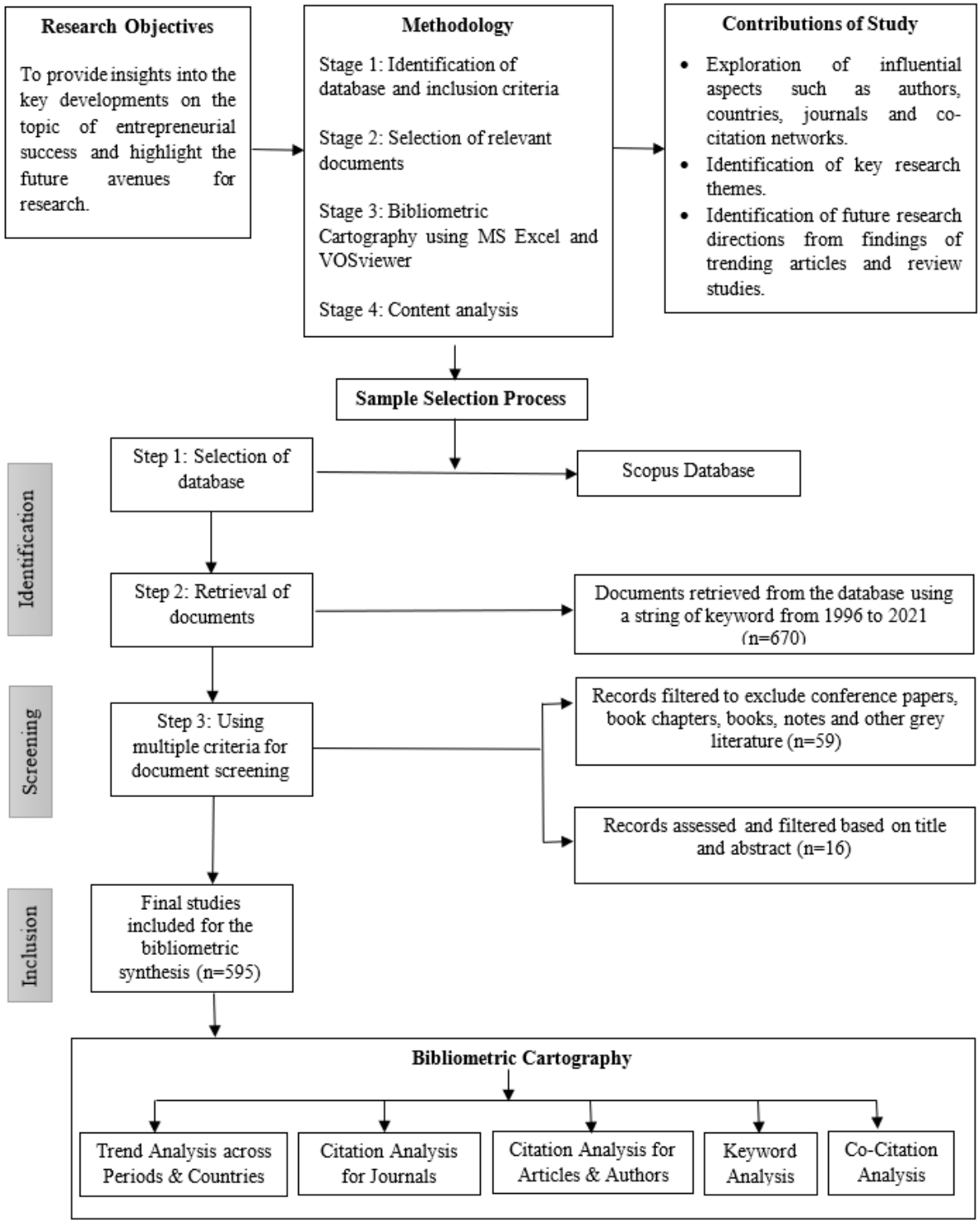
consequence of actions undertaken by entrepreneurs for the development of new and innovative products and services (Baker and Nelson, 2005). Dubin's theory showcases the phases or steps that can be used for theory or concept building (Dubin, 1978). It helps in the opportunity identification process and the development of the methodology which can help the entrepreneurs capture the prevailing opportunities in the market. The process comprises eight phases, where the first five phases are focused on developing the structure, and the last three phases are meant for empirical validation (Ardichvili et al, 2003). Lastly, Schumpeter's theory of economic development considers the significant role of the entrepreneurs and the innovations conducted by them in the process of economic development (Croitoru, 2012). It is a well-recognized fact that innovation and new technologies are not uniformly adopted by competing entrepreneurs (Jambulingam, 2018). The theory also accords special significance to the entrepreneur as a driver of change through constant innovation, which overcomes the obsolescence of past products (Schumpeter, & Backhaus, 2003). These theories have been critically analyzed in the cluster analysis section, along with highlighting other emerging yet prominent theories.

3. RESEARCH METHODOLOGY

The topic of entrepreneurial success has not been subjected to a quantitative review. This study fills this gap through a mixed-method study using a combination of bibliometric analysis, a quantitative method, and content analysis, which is a qualitative method. Bibliometric analysis has gained wide attention in recent years due to the ease of accessibility of open access software such as VOSviewer, R, and Gephi, as well as the cross-pollination of the themes from one domain to the other (Dhonthu et al., 2021). Scholars use bibliometric analysis for gaining knowledge about emerging trends, the performance of journals, clarity regarding research patterns, core journals, countries, and influential authors as well as publications (Singh et al., 2021a; Sardana and Singhania, 2022). This technique helps in producing a high impact on research by handling large volumes of statistical data (Dhonthu et al., 2021). Through scientific mapping, data has been analyzed to synthesize the existing research, giving a better clarification of the topic and field of study (Singhania et al., 2022a).

The first step of the technique involved the identification and selection of a database for the retrieval of documents. We used the Scopus database by Elsevier as it is known for its wide range of publications in social sciences, especially since 1996 (Vieira and Gomes, 2009). As a second step, we framed a string of keywords based on the literature review: "entrepreneur* success*" OR "entrepreneur* victory" OR "entrepreneur* achieve*". This led to the retrieval of 670 documents. Since the Scopus database is known to have a weak coverage of pre-1996 publications in the domain of social sciences (Harzing and Alakangas, 2016), the set of documents used for analysis was restricted to the period 1996 to 2021. As a third step, the screening of records was undertaken, which led to the filtration and elimination of 59 documents, including conference papers, book chapters, books, and notes to keep only journal articles in the English language (excluding articles in all other languages). Then, the records were assessed based on the title and abstract for eligibility. This further omitted 16 documents, leaving a sample of 595 documents that were finally used for the bibliometric synthesis. Figure 1 provides an overview of these steps along with the structural flow of the paper.

Figure 1. Structural flow of the paper



Source: The authors.

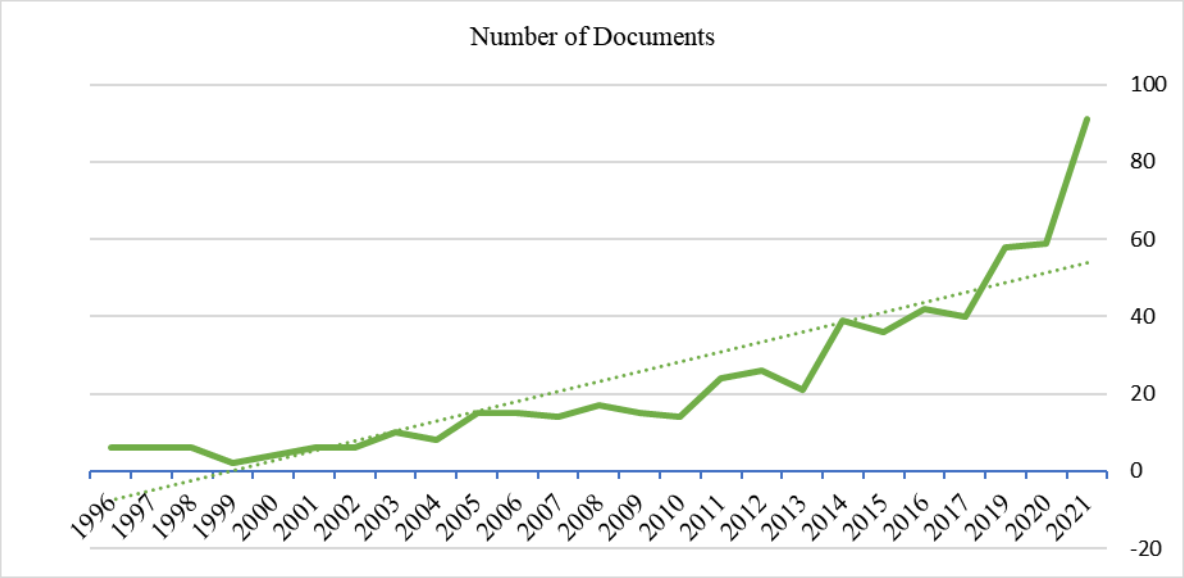
4. RESEARCH ANALYSIS

The study conducts bibliometric and topographical analyses of the research articles sourced from the Scopus database. In order to identify the pattern of development within the database, descriptive statistics are first performed using MS Excel to construct various tables, graphs, etc. Subsequently, to determine the relationship between the study items, VOSviewer (Van-Eck & Waltman, 2010) is employed for analysis, network diagrams, and science mapping (Baker et al., 2021).

4.1 Volume and Growth Trend of Published Studies

Our first analysis is focused on the volume of publications on entrepreneurial success, over the past 26 years, from 1996 to 2021. The analysis of 595 articles indicates that the knowledge base of entrepreneurial success has witnessed slow but steady growth, and there is wide potential for evolution in this area (Figure 2). Though at first glance the number of publications might seem low compared to the extended horizon of 26 years, we must also consider that this number does not cover a fair number of documents such as books, book chapters, conference papers, reports, etc. Figure 2 also highlights how the studies in the said domain have seen a steep rise post-2013. This persistent growth could be attributed to the initiatives undertaken by the government of various economies at the global level to push the start-up culture (Yin et al., 2019; Garg and Gupta, 2022). Countries have started realizing the potential of the start-up ecosystem, and research studies in the domain of entrepreneurial success act as game changers for boosting the initiatives taken by various government bodies. Moreover, the recent COVID-19 pandemic, which led to a massive loss of employment, also acted as a driver for self-employment and entrepreneurial initiatives (Dang and Nguyen, 2021).

Figure 2. Growth of research on entrepreneurial success over the past 26 years (1996-2021)



Source: Authors' creation based on Scopus database.

4.2 Publication Analysis across Countries

Table 1 presents the top 20 countries contributing to the topic of entrepreneurial success based on the number of citations of published documents. With the highest number of documents (197) and citations (8464), the United States of America (USA) proves to be the most productive. The authors of the documents published in the USA have been exploring the topics of success among small and medium enterprises (Eggers, 2020), women entrepreneurship (Bullough et al., 2022), and entrepreneurship education (Liguori and Winkler, 2020). The second position is taken by Germany, with 74 documents and 1955 citations. The third spot is occupied by the Netherlands while the United Kingdom (UK) assumes the fourth spot based on the number of citations. It is also interesting to note that though Sweden is found in sixth position in terms of the number of citations, it has the highest average citations per document (ACPD), thereby reflecting a higher impact in relative terms, vis-à-vis other countries. The above observations follow because the economic growth of a nation is positively correlated with the entrepreneurial activities taking place in the country (Stel et al., 2005). Factors such as innovation, employment, and productivity growth are driven by growing entrepreneurial activity (World Economic Forum, 2013). The USA has consistently seen upward trends, both in early-stage entrepreneurship as well as established entrepreneurial ventures (Global Entrepreneurship Monitor [GEM], 2019). With this comes the responsibility of research and publications across the country, which helps in innovation and development (Distanont and Khongmalai, 2020). It is for this reason that the USA tops the list of 20 countries.

Table 1. Top 20 most influential countries based on citation analysis

Rank	Country	No. of Documents	No. of Citations	ACPD	Type of Economy
1	United States of America	197	8464	42.96	Developed
2	Germany	74	1955	26.42	Developed
3	Netherlands	27	1748	64.74	Developed
4	United Kingdom	52	1063	20.44	Developed
5	Spain	26	702	27	Developed
6	Sweden	8	601	75.13	Developed
7	Australia	30	570	19	Developed
8	Taiwan	11	354	32.18	Developed
9	France	22	353	16.05	Developed
10	Norway	13	348	26.77	Developed
11	Canada	15	321	21.4	Developed
12	Switzerland	13	227	17.46	Developed
13	India	36	212	5.89	Developing
14	Poland	18	196	10.89	Developed
15	Hong Kong	5	180	36	Developed
16	Austria	7	134	19.14	Developed
17	China	41	127	3.1	Developing
18	New Zealand	9	126	14	Developed
19	United Arab Emirates	10	125	12.5	Developed
20	Italy	17	121	7.12	Developed

Source: Authors' creation based on Scopus database and reviews.

Note: ACPD= Average Citation Per Document

An important aspect is that 90% of entrepreneurial research has been undertaken in developed countries across the world. Though not a lot of developing countries engage in entrepreneurial research, countries like India and China have been able to capture a spot in the list of top 20 countries. India stood at 13th place with an ACPD of 5.89 and China stood at 17th place with an ACPD of 3.1. These two developing economies had the lowest ACPD amongst the list of countries mentioned in the table. On the other hand, developed countries like Sweden and Netherlands are found to have an ACPD of 75.31 and 64.74 respectively, which is the highest in the table. This indicates the difference in the impact of the research among developed and developing economies.

4.3 Citation Analysis for Journals

The Scopus database comprises several journals that publish entrepreneurial research. Table 2 presents the top 20 journals with the highest citations in the Scopus database. The *Journal of Business Venturing* by Elsevier, an A* listed journal, proved to be the most influential source, with 5638 citations and an ACPD of 352.38. The second position is *Small Business Economics* by Springer Nature with 793 citations and an ACPD of 66.08 followed by *Journal of Small Business and Enterprise Development* by Emerald with 354 citations and 59 ACPD. *Academy of Entrepreneurship Journal* and *Journal of Entrepreneurship Education*, earlier published by Allied Business Academies, appear in the top 20 list, but they were delisted from Scopus in 2021 and 2019, respectively.

Table 2: Top 20 most influential journals in the domain of entrepreneurial success, based on the citation analysis

Rank	Source	NOD	No. of Citations	ACPD	Cite Score	Scopus Coverage	ABDC listed	Publisher
1	Journal of Business Venturing	16	5638	352.38	14.6	1985 to Present	A*	Elsevier
2	Small Business Economics	12	793	66.08	10.7	1989 to Present	A	Springer Nature
3	Journal of Small Business and Enterprise Development	6	354	59	5.4	1994 to Present	C	Emerald
4	Entrepreneurship: Theory and Practice	6	313	52.17	13.7	2004 to Present	A*	Wiley-Blackwell
5	Entrepreneurship and Regional Development	8	282	35.25	8	1989 to Present	A	Taylor & Francis
6	Journal of Business Research	8	262	32.75	11.2	1973 to Present	A	Elsevier
7	International Journal of Entrepreneurial Behaviour and Research	10	245	24.5	8	1995 to Present	B	Emerald
8	Business Horizons	5	183	36.6	14	1957 to Present	C	Elsevier
9	Management Decision	7	166	23.71	7.9	1967 to Present	B	Emerald
10	Journal of Small Business Management	8	160	20	8.4	1996 to Present	A	Taylor & Francis
11	International Entrepreneurship and Management Journal	7	126	18	9.7	2006 to Present	C	Springer Nature
12	Journal of Small Business and Entrepreneurship	8	106	13.25	5	1996 to 1998; 2000 to 2014; 2016 to Present	C	Taylor & Francis
13	Journal of Developmental Entrepreneurship	7	102	14.57	2	2008 to Present	C	World Scientific
14	Academy of Entrepreneurship Journal	11	75	6.82	1.1	2009 to 2021	N.A.	Allied Business Academies
15	International Journal of Entrepreneurship and Small Business	10	69	6.9	1.3	2004 to Present	C	Inderscience Publishers
16	Journal of Entrepreneurship in Emerging Economies	7	61	8.71	4.6	2014 to Present	N.A.	Emerald
17	Frontiers in Psychology	12	38	3.17	4	2010 to Present	N.A.	Frontiers Media S.A.
18	Journal of Entrepreneurship Education	6	32	5.33	2.7	2009 to 2019	N.A.	Allied Business Academies
19	Business History	5	11	2.2	1.9	1958 to Present	A	Taylor & Francis
20	New England Journal of Entrepreneurship	5	8	1.6	2.6	1998 to Present	C	Emerald

Source: Authors' creation based on Scopus database and reviews.

Note: NOD= Number of Documents; ACPD= Average Citation Per Document

It can also be observed that business journals cover most of the entrepreneurial success research. Other journals publishing on entrepreneurship cater to areas such as management, economics, psychology, education, and history. The scholars working in this domain should note an interesting aspect: that despite being in the top 3 list based on the volume of publications, the *Frontiers in Psychology Journal* (with 12 published documents) is performing low in terms of the ACPD. This means that though enough papers are being published, the publications have not been able to generate substantial impact relative to publications in other journals. As pointed out before, entrepreneurship is one of the major contributors to the development of the economy (Doran et al., 2018). Beyond the economic factors, entrepreneurial success is influenced by psychological factors. Psychology is the foundation that evaluates the ability of an entrepreneur to recognize business opportunities and use them effectively to start and build a firm (Ramos-Rodrıguez et al., 2010). Similarly, different domains have different purposes in the journey of entrepreneurship, which are represented systematically by various journals.

4.4 Influential Authors and Articles

The study not only identifies the top countries and journals working in the field of entrepreneurship but also gives attention to the contributors as well as the most influential articles that have been published and have proved to be beneficial for academicians and entrepreneurs.

Table 3 lists the top 20 authors who have been contributing their efforts towards the said topic, ranked based on citation analysis. Michael Frese has proved to be the most influential author, with 1209 citations and a

h-index of 112. Some of his research articles such as "*Human Capital and Entrepreneurial Success: A Meta-Analytical Review"* and "*Towards a Psychology of Entrepreneurship- An Action Theory Perspective",* have made it to the top 10 most influential articles in this domain. The second most impactful author is Eva Schmitt-Rodermund, with 285 citations and a h-index of 28. The citations received by Michael Frese, as well as his h-index are more than four times those of Eva Schmitt-Rodermund. It is noteworthy that most of the top authors belong to Germany and other parts of Europe. Though the USA topped the list of countries contributing to research on entrepreneurial success, none of the top authors belong to the USA. This shows that the country is providing a substantial quantity of research but lacks impactful authors working in this area, vis-a-vis countries in Europe. Also, countries such as India and China, which stood at 13th and 17th place respectively in the list of top 20 countries and are the only two developing countries that were able to find a spot in the list, have been able to find a place even in the list of most influential authors. Jianxin Zhang (#12) belongs to China and Ranbir Singh (#20) belongs to India.

Table 3: Top 20 authors publishing in the domain of entrepreneurial success, based on citation analysis

Rank	Author's Name	NOD	Citations	ACPD	Country Affiliation	Affiliation	h-index
1	Michael Frese	6	1209	201.5	Germany	University of Lueneburg	112
2	Eva Schmitt-Rodermund	3	285	95	Germany	University of Applied Sciences Potsdam	28
3	Yu-Yu Chang	4	172	43	Taiwan	National Cheng Kung University	12
4	Chen, Ming-Huei	4	172	43	Taiwan	National Chung Hsing University	27
5	Rebecca Fisher	3	127	42.33	England	Coventry University	2
6	Ute Stephan	3	122	40.67	England	King's College London	34
7	Martin Obschonka	4	119	29.75	Netherlands	University of Amsterdam	41
8	Rainer K. Silbereisen	4	119	29.75	Germany	Friedrich Schiller Universität Jena	69
9	Marcin Waldemar Staniewski	6	101	16.83	Poland	University of Economics and Human Sciences in Warsaw	16
10	Katarzyna Awruk	5	58	11.6	Poland	Akademia Ekonomiczno-Humanistyczna w Warszawie	8
11	Josette Dijkhuizen	4	38	9.5	Netherlands	Maastricht School of Management	N.A.
12	Jianxin Zhang	3	30	10	China	N.A.	12
13	Roy Thurik	3	27	9	France and Netherlands	Montpellier Business School and Erasmus School of Economics	99
14	Alexandros Kakouris	3	24	8	Greece	University of Peloponnese	13
15	Agnieszka Kurczewska	3	13	4.33	Poland	University of Lodz	15
16	Christian Leyh	4	10	2.5	Germany	Fraunhofer Center for International Management and Knowledge Economy	12
17	André Krischke	3	5	1.67	Germany	Hochschule München	N.A.
18	Michał Mackiewicz	3	5	1.67	Poland	University of Lodz	11
19	Susanne Strahringer	3	5	1.67	Germany	Technische Universität Dresden	11
20	Ranbir Singh	4	2	0.5	India	N.A.	2

Source: Authors' creation based on Scopus database and reviews.

Note: NOD= Number of Documents; ACPD= Average Citations Per Document

As for the top articles, Table 4 presents the top 10 most influential articles which have been most cited (according to the Scopus database) and contributed significantly to the field of our study. Each of the articles has received more than 150 citations. Out of these, the latest article was published in 2013 and underlines the linkages between entrepreneurial success and gender diversity. These influential research articles can be categorized into four broad domains: the role of psychological factors in entrepreneurial success; the role of gender; the impact of family business; and other factors affecting success in entrepreneurship. Psychology is one of the most pertinent factors with respect to entrepreneurship as most of the articles talk about the role of psychology in entrepreneurship. Psychology is a mechanism that helps entrepreneurs in coping with difficult and challenging situations (Zou et al., 2015). Most of these articles are published in ABDC-listed journals and

contribute significantly to the literature. Table 5 also reiterates that the *Journal of Business Venturing* is the top contributing journal as 50% of the top 10 articles have been published in this journal.

Table 4: Top 10 most influential articles in the field of entrepreneurial success, based on citation analysis

Rank	Document Name	Journal Name	Citations	Year	Type of Paper
1	Linking two dimensions of entrepreneurial orientation to firm performance: The moderating role of environment and industry life cycle	Journal of Business Venturing	1527	2001	Empirical
2	Does entrepreneurial self-efficacy distinguish entrepreneurs from managers?	Journal of Business Venturing	1435	1998	Empirical
3	Human capital and entrepreneurial success: A meta-analytical review	Journal of Business Venturing	832	2011	Empirical
4	Networks in entrepreneurship: The case of high-technology firms	Small Business Economics	491	2003	Empirical
5	Entrepreneurship and female youth: knowledge, attitudes, gender differences, and educational practices	Journal of Business Venturing	245	1998	Empirical
6	Business formation: A network approach	Scandinavian Journal of Management	213	1988	Conceptual
7	Linking family-to-business enrichment and support to entrepreneurial success: Do female and male entrepreneurs experience different outcomes?	Journal of Business Venturing	205	2013	Empirical
8	Pathways to successful entrepreneurship: Parenting, personality, early entrepreneurial competence, and interests	Journal of Vocational Behavior	190	2004	Empirical
9	Towards a psychology of entrepreneurship: An action theory perspective	Foundations and Trends in Entrepreneurship	181	2009	Conceptual
10	Psychological perspectives on entrepreneurship: Cognitive and social factors in entrepreneurs' success	Current Directions in Psychological Science	154	2000	Conceptual

Source: Authors' creation based on Scopus database and reviews.

4.5 Keyword Analysis

We undertake a keyword analysis to identify the research themes studied predominantly in the entrepreneurial success literature. A novel technique of quantitative analysis, keyword analysis is a subfield of bibliometric review that has its roots in library and information science. It aims to identify repetitive words, trending domains, and emerging topics of research (Qian et al., 2019).

Table 5: Top 10 keywords occurring for the highest number of times

Rank	Keyword	No. of Occurrences
1	Entrepreneurs	24
2	Gender	21
3	Innovation	21
4	Women entrepreneurs	16
5	Human capital	15
6	Education	14
7	Social capital	14
8	Entrepreneurship education	13
9	Sustainable development	12
10	Entrepreneurial orientation	11

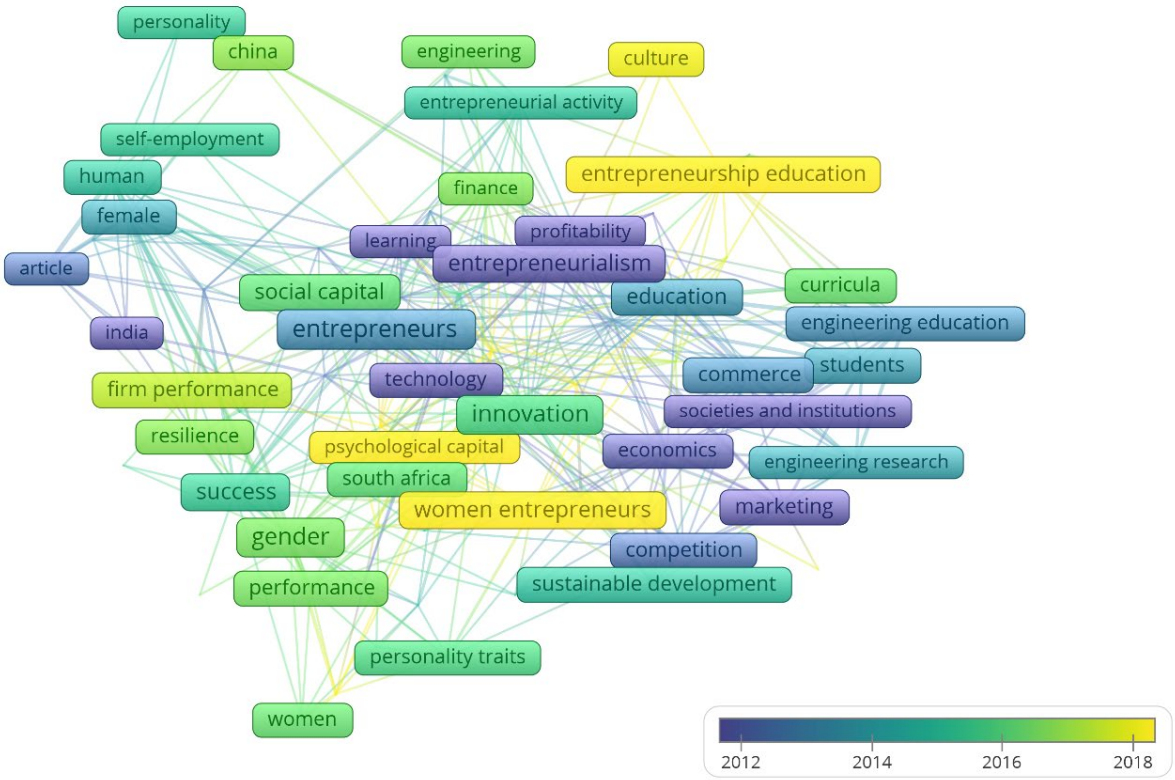
Source: Authors' compilation using Scopus database.

Table 5 shows the top keywords which have occurred in the database and these consist of topics like *Entrepreneurs, Gender, Innovation,* and *Women Entrepreneurs.* This highlights the major dimensions that have been studied in the past literature under the theme of entrepreneurial success. Gender-based entrepreneurship has gained momentum in recent years owing to a larger proportion of women entrepreneurs taking the lead in the market (Rosca et al., 2020). Similarly, innovation is the backbone of a successful entrepreneurial venture (Iriobe, 2018), and this is evident from the frequent usage of this term in the literature. The list continues with the terms *Human capital, Education, and Social Capital.* Human capital is the biggest asset of a firm, and works

as a tool for the success of a firm. It possesses a direct relationship with entrepreneurial success, both in new as well as established businesses (Elsafty et al., 2020). Education is an integral part of entrepreneurship as well. In the era of a changing entrepreneurial environment, education has become a need for current as well as future entrepreneurs. Investment and efforts in entrepreneurship education have increased in the past few years and are continuously witnessing an uprise (Bauman and Lucy, 2021). Social capital adds value by providing entrepreneurs with the right network and direction (Xie et al., 2021). Lastly, keywords like *Sustainable Development* have received attention due to climate destruction as well as pandemics (such as COVID-19) that have been impacting the world. Moreover, the emphasis of the United Nations (UN) on sustainable development goals has further given a necessary push to research in this domain (Singh et al., 2021b). The dimension of *Entrepreneurial Orientation* has also been looked at to understand the dynamics of top management style, organizational configuration, and new entry initiatives. The understanding of these aspects will help both the entrepreneurs as well as researchers to plan the multilevel conceptualization on a larger scale (Wales et al., 2020).

Additionally, keyword co-occurrence analysis has allowed us to locate emerging topics within the subject domain (Bhattacharya and Basu, 1998). The emerging topics have been represented by the colour yellow in Figure 4. The years after the global financial crisis witnessed the hefty usage of terms like technology (Ford et al., 2010), profitability (Kiviluoto, 2011), and economics (Brown and Thornton, 2013), and most of the entrepreneurial research revolved around these topics. Since entrepreneurship has gained wide popularity in recent times, the topics of research have also evolved to focus on the themes of psychological capital (Tang, 2020; Gao et al., 2020), entrepreneurship education (Fayolle, 2018; Kyrö, 2018), women entrepreneurs (Panda, 2018; Sajjad et al., 2020) and culture (Lounsbury, 2019). The list of the most influential articles shows that most of the top 10 articles consist of similar topics. Entrepreneurs' behaviour and characteristics are integral factors in evaluating the success of a venture. Gender and culture impact entrepreneurship and entrepreneurial success. These emerging topics have a great deal of scope for future research that can be explored by scholars working in the entrepreneurial success domain.

Figure 4: Temporal overlay on a keyword co-occurrence map for the entrepreneurial success literature



Source: Authors' creation based on Scopus database and VOS viewer.

Figure 4 also shows the network between the topics which are connected using various nodes. The connection between these key concepts indicates the existence of a relationship between them. The relationship among the key concepts is visible and has been discussed in the cluster analysis as well as in future research avenues.

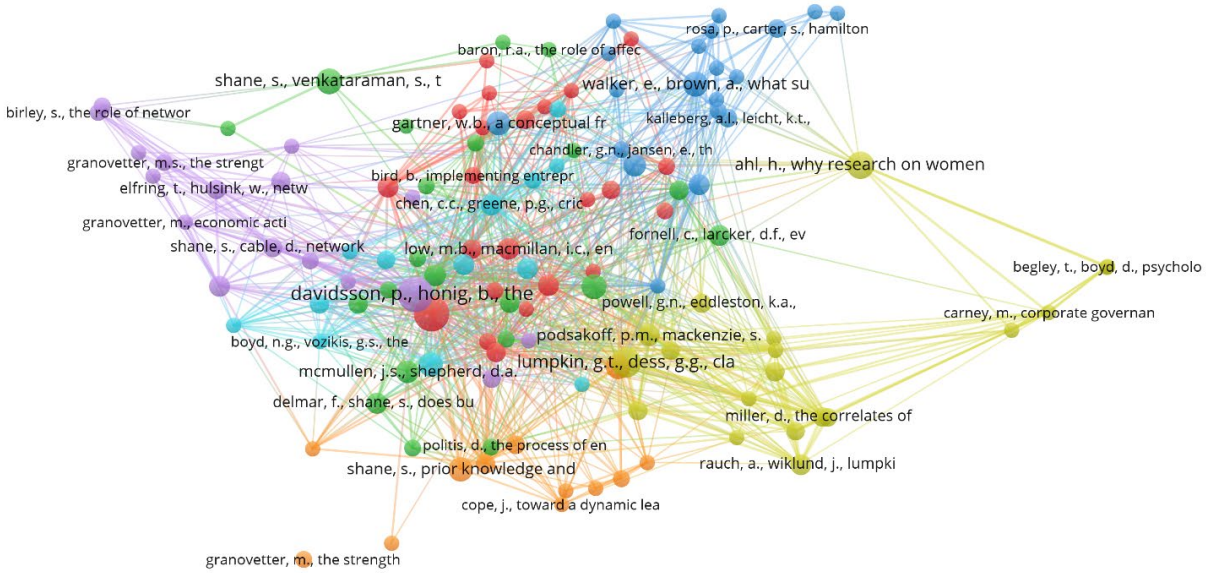
4.6 Cluster and Content Analysis Based on Co-Citation Analysis

Co-citation analysis is often used as a method to uncover the intellectual developments taking place in a particular field of research and identify emerging themes (Donthu et al., 2021). The method can be used to identify the major themes which will also be useful in determining the clusters and the existing inter-relationship (Bhaiswar et al., 2021) from the cited references.

Cluster analysis refers to the data analysis technique employed to identify the groups or segments which naturally occur together without any pre-defined set of rules. Due to the lack of any pre-defined data points, this method is often termed an unsupervised learning approach (Singhania et al., 2022; Peñas et al., 2023). The cluster analysis has been conducted using the co-citation method to identify the various sets of clusters in the domain of entrepreneurial success.

In the network map portrayed in Figure 5, the nodes are representative of the cited references, whereas the size of the nodes depicts the number of articles in which the document has been co-cited. Out of the 34656 cited references, 125 references meet the criterion of a minimum of 5 citations and have been included to visualize key thematic clusters in our research domain. Content analysis has been employed to understand the 6 themes emerging from these clusters.

Figure 5: Co-citation network based on cited references with minimum 5 citations



Source: Authors' creation using VOSviewer.

Cluster 1: Theories underpinning entrepreneurial success

Cluster 1 (red) is the largest cluster, with references emerging as early as 1934. The studies in this cluster discuss various theories involved in the identification and development of entrepreneurial opportunities, some of which are 'discovery theory', 'creation theory' (Alvarez and Barney, 2007), 'Dubin's theory' (Ardichvili et al., 2003), 'cognitive theory' (Palich and Bagby, 1995), and the 'theory of economic development' (Schumpeter, 1934). These theories have been discussed in detail in the section on 'Literature Review'. An entrepreneur's intention and state of mind, combined with experience and action toward a business concept, gives a direction for the organization at the inception stage, which further helps in survival, development, and growth (Bird, 1988). Some opportunities exist and some are waiting to be identified by the actions of an entrepreneur. Strategic entrepreneurship is a unique concept that involves the opportunity-seeking and advantage-seeking behaviour of an entrepreneur, which apparently results in the better performance of the firm (Ireland et al., 2003; Mazzei, 2018). It is stated that factors such as an entrepreneur's passion, perception of skills, abilities, luck, and value of the opportunity comprehended together lead to entrepreneurial effectiveness and persistence (Gatewood, 1995; Cardon, 2009; Kakarika et al., 2022).

Apart from the major theories highlighted by this cluster, certain theories are emerging in the entrepreneurship literature, and have the potential to affect it. These include the jack-of-all-trades theory, o-ring theory, the theory of entrepreneurial bricolage, and the theory of effectuation. The theory of Jack-of-all-trades was developed by Lazear (2004) and mathematically proved by Kaiser (2012). The theory is based on a balanced skills mix and assumes that to pursue entrepreneurship as a career option individuals need to possess sufficient skills and knowledge of a variety of fields (Wagner, 2003). Entrepreneurial success depends on the entrepreneurs' knowledge and experiences, which are accumulated by pursuing varied curricula. This helps in

building practical experiences and gaining insights from a range of functions that further helps in fostering entrepreneurial skills leading to entrepreneurial success (Kurczewska and Mackiewicz, 2020). As far as o-ring theory is concerned, it suggests the segregation of individuals according to their abilities, as different individuals supply different sets of talents and skills. This theory, seminally developed by Kremer (1993) and theoretically by Fabel (2004), believes in the power of teamwork and its importance in the success of a venture. The theory of entrepreneurial bricolage is a behaviour-based theory of entrepreneurship, proposed by the anthropologist Lévi-Strauss (1962). The theory tries to understand the behaviour of entrepreneurs to manage uncertainties, survive, and flourish even when faced with resource constraints (Senyard et al., 2009). It explores the possibility of the emergence of entrepreneurship with limited resources available at hand (Kickula et al., 2018). Finally, the theory of effectuation came into existence after the seminal works of Sarasvathy et al. (1998) and Sarasvathy (2001). According to Sarasvathy (2001), the future is unpredictable. From the logic of effectuation, it is better to control an unpredictable situation than to predict an uncertain one. Effectuation pre-estimates the affordable losses and lets entrepreneurs experiment with many strategies within limited means.

Cluster 2: Drivers of entrepreneurial success

This is the second largest cluster, depicted in green, with referenced documents emerging as early as 1981. One of the most important non-financial drivers of entrepreneurial success, that is psychology, is distinguished from the other behavioural sciences as it focuses more on the behaviour of an individual (Ajzen, 2020). In turn, behaviour is influenced by the dynamism of the external environment and an individual's personal choice (Shaver and Scott, 1992). Extending this concept to entrepreneurship, an entrepreneur's behaviour impacts many factors, such as readiness for innovation, proactive personality, generalized self-efficacy, stress tolerance, the need for autonomy, and locus of control, which have a significant correlation with business creation and business success (Brandstätter, 2011; Jawabri, 2020; Kurniawan et al., 2022). Research studies in this cluster have also studied the relationship between entrepreneurial traits and skill (passion, tenacity, and new resource skill) and situational specific motivation (communicated vision, self-efficacy, and goals) to further venture growth (Pathirana and Semasinghe, 2018; Rylková and Čemerková, 2020). Goals, self-efficacy, and communicated vision have direct effects on venture growth, and these factors mediate the effects of passion, tenacity, and new resource skill on subsequent growth (Baum and Locke, 2004). Prior knowledge of customer problems leads to the identification of more opportunities, and it also moderates the relationship between potential financial reward and opportunity identification (Shepherd and DeTienne, 2005).

The other important drivers of entrepreneurial success, as pointed out by cluster 2, can be categorized into economic drivers, psychological drivers, social drivers, and personal drivers. Traditionally, the profits of the firm have been considered solely as an indicator and driver of business performance as well as success. Though now there are various other factors involved in the evaluation of business success, economic performance still holds a very strong position in this evaluation (Perez and Canino, 2009). Psychological drivers comprise a) resilience, which is a desirable characteristic required by both the entrepreneurs as well as their ventures (Bullough et al., 2014; Baluku et al., 2016), b) self-efficacy (Goel and Karri, 2006; Baluku et al., 2016), since entrepreneurs with high self-efficacy identify as well as exploit the available opportunities and develop positive entrepreneurial intentions (Dimov, 2010), thereby harnessing the optimum levels of motivation and c) optimism, as entrepreneurs can take risks because of their confidence and optimistic attitude (Hmieleski and Baron, 2009) and it also has a significant positive relationship with the financial aspect of the firm (Baluku et al., 2016). The social drivers consist of a) workplace relationships, since having a healthy bond with customers and clients by providing them with loyalty and satisfaction also helps a firm in gaining long-term success (Wach et al., 2016), and b) societal impact (Tur-Porcar et al., 2018), which takes into account the ability of the entrepreneur to ensure that innovation and creativity are used to develop products and services with minimal and judicial application and exploitation of resources. The last set of drivers, that is personal drivers, comprises a) education, which is a basic level of knowledge and skills and which helps aspiring and budding entrepreneurs to identify and exploit market opportunities for profitable ventures (Barreneche, 2014), b) innovativeness and creativity (Somsuk and Laosirihongthong, 2014), both of which support the economic development of the country, work as a catalyst for employment generation as well as wealth creation, and lead to technological advancements, thereby creating more opportunities (Mazzucato, 2011) and c) personal aspects, including personality traits and characteristics of self-confidence, independence, autonomy, risk-taking propensity, persistence, networking, problem-solving ability, learning from failure, and using criticism (Rauch and Frese, 2007).

Cluster 3: Women and entrepreneurial success

Cluster 3, represented in blue, focuses on the journey of women entrepreneurs. The relationship between gender and business performance is complex, but gender is a key determinant of success, even after controlling for other factors (Rosa et al., 1996; Bednar et al., 2021; Olivas et al., 2022). Women entrepreneurs face problems when they enter the business, but factors such as family support, social ties, and internal motivation add positively and significantly to the success of women entrepreneurs (Alam et al., 2011; Panda, 2018). Work-

life balance too plays a significant role in the evaluation and success of women entrepreneurs (Davidsson, 1991; Gorgievski et al., 2011). The entrepreneurial success of women entrepreneurs is also majorly attributed to internal factors such as the need for achievement, risk-taking, and self-confidence (Khan et al., 2021). These factors accord much-needed motivation to women entrepreneurs. On the other hand, factors such as the socio-cultural status of women, difficulty in assessing the network of information, and problems in obtaining funds are the major factors that limit the performance of firms led by women entrepreneurs (Hasaan et. al., 2017). There is also a plethora of studies that contend that gender diversity or gender-related aspects do not impact firm performance. In a recent study by Singh et al. (2019), it was argued that having a gender-diverse board does not impact the performance of recently listed IPO (Initial Public Offer) firms. Similarly, another study by Singhania et al. (2022b) identified that having gender-based aspects on various committees does not impact accounting-based performance. Thus, this aspect as to whether gender-based elements impact performance is still debatable and can be explored in the future (Singhania et al., 2023a; Singhania et al., 2023b; Singh et al., 2023).

Cluster 4: Entrepreneurial orientation and entrepreneurial success

The fourth cluster is represented in yellow. The term entrepreneurial orientation (EO) has been used to refer to the strategy-making process and the style of firms that engage in entrepreneurial activities (Lumpkin et al., 2001). A popular model of EO suggests that it has five encompassing dimensions: autonomy, innovativeness, risk-taking, proactiveness, and competitive aggressiveness (Lumpkin et al., 1996; Al-Mamary and Alshallaqi, 2022). Entrepreneurial orientation shows a direct correlation not only with the success and performance of existing entrepreneurial firms but also with the survival and growth of new entrepreneurial ventures (Rauch et al., 2008; Stam and Elfring, 2008). Business performance is also determined by the psychological attributes of the entrepreneur such as the need for achievement, locus of control, risk-taking propensity, and tolerance of ambiguity (Ndofirepi, 2020). These are regarded as hallmarks of an entrepreneurial personality (Begley et al., 1987).

The theme of entrepreneurial orientation has often been linked to the concept of international entrepreneurship, which encompasses the innovativeness and risk-taking abilities that transcend national boundaries for value creation (McDougall and Oviatt, 2000). However, the relationship is indirect, and it has been found that international entrepreneurship mediates the role between entrepreneurial orientation and entrepreneurial success. The firms which pivot and take advantage of international moves at the right moment are bound to gain entrepreneurial success (Kollmann and Christofor, 2014).

Cluster 5: Resources and entrepreneurial success

This cluster (purple) has an integral role to play in the entrepreneurial journey as the documents in this cluster throw light on the way entrepreneurs interact with the networks in their local environment during the process of starting a new firm (Cavallo et al., 2019). It also discusses the development of entrepreneurship in a resource-constrained environment. It has been found that an entrepreneur is not only seeking the resources of equipment, space, and money, but also advice, information, and reassurance (Birley, 2000; Slavec and Prodan, 2012). Consequently, the help and guidance received from both the formal networks (banks, accountants, lawyers) and the informal networks (family, friends, business contacts) will influence the nature of the firm substantially (Birley, 1985; Watson, 2012). Network resources, networking activities, and network support are heavily used to establish new firms (Malecki, 2018). Further, entrepreneurs who can refer to a broad and diverse social network and who receive a great deal of support from their network tend to be more successful (Brüderl and Preisendörfer, 1998). It has also been found that entrepreneurs in resource-poor environments are able to render unique services by recombining the elements at hand for new purposes (Archer et al., 2022). They believe in making do with what is at hand (Baker and Nelson, 2005). Past literature clearly documents a direct nexus between resources, dynamic capabilities, and entrepreneurial success (Lu et al., 2010, Hitt et al., 2011). The resources can be categorized into various types such as tangible and intangible assets, financial resources, techniques, manpower, social networks, and goodwill (Huang, 2016). These resources act as a catalyst and with the presence of these resources, the speed of entrepreneurial success is enhanced (Huang, 2016). Along similar lines, dynamic capabilities are those high-level skills that integrate, reorganize, and configure the existing resources to improve the performance of the firms leading to entrepreneurial success (Raza et al., 2018).

Cluster 6: Cognitive learning and entrepreneurial success

This cluster represents two overlapping themes (light blue and orange). It discusses the learning attitude of an entrepreneur (Eliyana et al., 2020) and emphasizes a framework that identifies three main components in the process of entrepreneurial learning: entrepreneurs' career experience, the transformation process, and entrepreneurial knowledge in terms of effectiveness in recognizing and acting on entrepreneurial opportunities as well as coping with the liabilities of newness (Politis, 2005). An entrepreneur needs to possess a dynamic learning potential with three distinctive but interrelated elements of entrepreneurial learning, which are dynamic temporal phases, interrelated processes, and overarching characteristics (Cope, 2005). Apart from learning

attitude, self-efficacy, and achievement motivation are significantly correlated with both the choice of an entrepreneurial career and entrepreneurial performance (Collins et al., 2004; Staniewski and Awruk, 2019). Self-efficacy is the confidence of an entrepreneur that he/she is capable of various roles and tasks, which include marketing, innovation, management, risk-taking, and financial control (Chen et al., 1998; Bakkar et al., 2021). The entrepreneur's behaviour, such as the need for achievement, innovativeness, stress tolerance, need for autonomy, and proactive personality matches positively with the entrepreneur's objective of business creation and business success (Rauch and Frese, 2007).

The literature in this cluster also suggests that in firms led by entrepreneurs with a high degree of absorptive capacity, the chances of entrepreneurial success get amplified (Khan et al., 2020). Absorptive capacity refers to the firm's ability to generate innovation and facilitate learning (Cohen and Levinthal, 1990). Similarly, learning capabilities refer to the aggregate of skills that can help firms leverage competitive advantage in the markets through innovation (Barão et al., 2017), foster the growth of the firms, and equip them for market survival in the long run (Chang et al., 2012). Furthermore, these skills are the core aspects that help firms in effective management and lead to entrepreneurial success as well as improved performance (Khan et al., 2021).

5. DISCUSSION AND FUTURE RESEARCH AVENUES

The current study has answered the research questions significantly using bibliometric analysis as well as content analysis and has provided an overview of the topic of entrepreneurial success. The volume analysis has consistently shown an upward trend (RQ_1), starting from 6 documents in 1996 to a significant number of 91 documents in 2021. The number of documents from the year 2020 to 2021 showed the biggest jump. One of the major reasons for this was the need to capture the impact of the COVID-19 pandemic by researchers, which severely impacted almost every sector of the economy negatively, but had a positive impact on entrepreneurship (Sharma et al., 2022). The growing trend of publications shows that the research in this arena has not yet been saturated, and, therefore, future research scholars should continue to undertake studies that have implications for policymaking as well as practice. In terms of geographical distribution, the USA tops the list with 197 documents and leads both in quantity and quality of research, followed by the European countries of Germany, Netherlands, and the UK. These economies, being developed economies, understand the importance of entrepreneurship, and apparently provide several benefits and incentives encouraging people to contribute towards entrepreneurship (Distanont and Khongmalai, 2020). However, with a large number of startups and unicorns being recognized in developing countries as well, it is high time to undertake research on various entrepreneurial aspects through cross-country and multi-country studies that focus on emerging economies.

Considering the journals contributing towards entrepreneurship (RQ_2), the top position lies with the *Journal of Business Venturing,* which is a Scopus-listed, A* journal published by Elsevier. The high number of citations and ACPD describes the quality as well as the impact of the research published in this journal. Researchers in this area should gain insights from reading or publishing in such impactful journals. The most influential authors (RQ_3) in this domain belong to developed countries, but some developing countries like India and China have also been able to capture a spot. Michael Frese from Germany has contributed immensely to this field, with 6 documents and 1209 citations. The contributions of the top authors are highly valuable as they lay down the foundation for the development of various new concepts and enable cross-country studies on entrepreneurship. However, there is also a need for more collaboration among authors from different countries, affiliations, and backgrounds, in order to promote the integration of ideas as well as expertise. As far as the influential research documents are concerned (RQ_4), the 10 most cited papers have been published within the year range of 1988 to 2013. Two out of ten papers have been authored by Michael Frese and most of the papers are empirical in nature, which has helped in shaping the present research. Some of these impactful research articles build the basis for seminal work in the literature, and hence, are a must-read for academics as well as practitioners.

The evolution of the topics of research on entrepreneurial success (RQ_5) over the years highlights the interest of researchers in this domain and the dynamic nature of the topics. The cluster analysis conducted with the help of the co-citation method highlighted six prominent clusters, including the theories of entrepreneurial success, the drivers and factors, the role of women entrepreneurs, the significance of entrepreneurial orientation, the use of resources, and the importance of cognitive learning in the entrepreneurship literature. An insight into each of these major themes would shape the way a country encourages, supports, and facilitates the activities of entrepreneurial ventures.

In the early 2010s, the keywords that were frequently used were 'profitability', 'learning', 'technology', and 'marketing'. These topics still hold a relevant position in this field, both theoretically as well as practically, but after considering the dynamism of the environment, new topics have emerged (RQ_6), such as 'psychological capital', 'culture', 'entrepreneurial education', and 'women entrepreneurs'. Psychological capital is now being recognized as one of the most important assets of an entrepreneur as it enables them to keep going despite failures or rejection. Culture is another aspect that cannot be ignored by anyone who wants to innovate and run a

successful business venture, and, hence, research studies incorporating the understanding of culture in the success of a firm will soon take center stage. With many institutions now offering a course in entrepreneurship, and students keen on learning the requisite skills, the ability of entrepreneurial education to add relevance to the entrepreneurial culture also needs to be tested in various contexts. Finally, given the increasing role of women in the workforce and with them breaking the barriers, women's entrepreneurship is now becoming a separate area of study. Ideas based on these recently trending topics can be considered by scholars to fill the void in research. Moreover, these emerging themes will provide new dimensions for exploration in the field of entrepreneurial success.

5.2 Future Research Avenues

Entrepreneurial research is still emerging, because of which the scope of research in this field is immense. Some of the notable themes that could be turned into future research areas, as highlighted by the keyword analysis and the literature review, are as follows:

1) Impact of Psychological Capital

Psychological capital helps to examine the ability of the entrepreneur to recognize as well as exploit business opportunities (Ramos-Rodrıguez et al., 2010) and it, therefore, plays an essential role in the success of entrepreneurship (Baluku et al., 2016). Psychological capital comprises four elements: hope, efficacy, resilience, and optimism (Luthans et al., 2007). These factors work together inseparably and support an entrepreneur emotionally as well as mentally during stressful and pressurizing situations, leading to entrepreneurial success. The theme of entrepreneurship has seen dynamic changes where young college students are entering entrepreneurial ventures. Given this context, it would be fruitful to study the impact of various other emerging components of psychological capital, such as the role of self-efficacy (Maslakçı et al., 2022), the role of the personality traits and entrepreneurial intentions (Vaiz, and Ekemen) and diversity brought in by gender (Maslakçı et al., 2022), on the entrepreneurial success.

2) Studies on Entrepreneurship Education

With the rise in entrepreneurial activity, education in entrepreneurship plays an important role as the entrepreneurship failure rate is quite high (Cantamessa et al., 2018). An entrepreneur requires a basic set of skills and knowledge in various domains. The impact of education on entrepreneurial success is potentially strong as it helps an entrepreneur acquire and apply the diverse skills required to compete and be successful (Kolstad and Wiig, 2015). However, various aspects such as the role of teaching effectiveness in entrepreneurship (Liu et al., 2022), the role of social capital in building entrepreneurship education (Putro et al., 2022), the impact of entrepreneurship education on entrepreneurial creativity (Wang et al., 2022) are themes which require further exploration.

3) The Success of Women Entrepreneurs

Promoting women's entrepreneurship not only contributes toward economic development but the social development of the country as well (Jamali, 2009). The rate of growth of women entrepreneurs has witnessed an uprise and has positively impacted overall welfare and consumption (Minniti and Naudé, 2010). Earlier, entrepreneurship was considered a man’s domain (Jennings and Brush, 2013). But now, we are moving towards a more equitable world, with opportunities available to everyone without any discrimination among the genders. Moreover, the motivation and challenges of being a women entrepreneur vary significantly across nations (Rastogi et al., 2022). Nevertheless, raising finance as women entrepreneurs is another set of dynamic issues (Shaikh et.al., 2022) which needs to be addressed to promote more and more women entrepreneurs.

4) Culture-Specific Studies

Each nation possesses a different culture, and an entrepreneur is obliged to respect and follow the country’s culture to grow and survive. Numerous research studies have taken place concerning culture and entrepreneurship over the years (Lounsbury et al., 2019; Audretsch, 2020). It is difficult to make generalizations about different nations that have different cultures or levels of economic development (Siu and Lo, 2013), but at the same time, cross-cultural innovation (Stephan, 2022) can prove to be significant from the market’s point of view if explored in the right direction. Recent literature also points towards the importance of adequate training in building a corporate culture (González-Tejero and Molina, 2022), which can push entrepreneurial ventures for small and medium enterprises (SMEs). Therefore, an entrepreneur needs to be aware of a country’s culture before taking any decisions, while at the same time, policies made by a country should cater to entrepreneurial activities with the country’s culture in mind (Uddin and Akther, 2019). There is still a lot of scope in this topic as culture is a sensitive topic that has the feeling of society attached to it.

This study has theoretical as well as practical implications as the outcomes can be used not only by future scholars but also by budding entrepreneurs in searching for gaps and opportunities and taking suitable actions on

the same. It will help the existing entrepreneurial ventures with better decision-making as well as provide early access to future opportunities and the detection of threats. The literature talks about the importance of entrepreneurial education, which is equally or even more important compared to other subjects. If taken care of, it can lead to a rise in entrepreneurial activities, thereby enabling the economic growth of the country. Thus, entrepreneurship, if explored effectively, can prove to be one of the most effective tools for the economic growth and development of a nation.

6. CONCLUSION

This study has relied on bibliometric analysis, combined with content analysis, to understand the domain of entrepreneurial success in a detailed manner and elaborate on the main trends, themes, and research trajectories of this field of study (Gao et al., 2021).

The study highlighted the various countries leading in terms of publications on entrepreneurial success, some of which are the USA, Germany, Netherlands, and the UK. The list consisted of both developed as well as developing countries, but the developed countries dominated the list, with around 90% of the publications. Subsequently, the study analyzed top journals publishing on entrepreneurship and related topics. The list consisted of 40% journals from business domains and the rest from other disciplines like management, economics, psychology, education, and history. As far as the author's contributions are concerned, the most influential authors mainly belong to Germany and parts of Europe, which sheds light on the quality of research and research opportunities in Europe. It also shows that the *Journal of Business Venturing* is the top journal publishing articles related to entrepreneurship, as 50% of the top articles were published in the same journal. The most influential articles broadly talk about four domains, which are psychology, gender, family, and success. The study also explored the evolution of entrepreneurial research topics over the years, followed by the identification of 6 clusters, which were examined further through content analysis. Future research agendas are elaborated upon based on the emerging topics and themes of study in the entrepreneurial success literature.

Although the analysis consists of all the relevant research questions required for the scholars working on the topic, it suffers from certain limitations. The study relies on the Scopus citation database only, because of which some studies, which are beyond the coverage of the Scopus database, might have been overlooked. Another point to be considered is that the bibliometric analysis may favour old researchers (Wang et al., 2017), due to its tendency to research by data rather than judgement. But it is noteworthy that a high number of citations do not represent a high impact on scientific results (Drew et al., 2016).

REFERENCES

Abu, F., Yunus, A. R., Majid, I. A., Jabar, J., Aris, A., Sakidin, H., & Ahmad, A. (2014). Technology Acceptance Model (TAM): Empowering smart customer to participate in electricity supply system. Journal of Technology Management and Technopreneurship (JTMT), 2(1). Retrieved from https://jtmt.utem.edu.my/jtmt/article/view/59

Agarwal, R., & Dahm, M. J. (2015). Success factors in independent ethnic restaurants. Journal of Foodservice Business Research, 18(1), 20-33. https://doi.org/10.1080/15378020.2015.995749

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. Human Behavior and Emerging Technologies, 2(4), 314-324. https://doi.org/10.1002/hbe2.195

Alam, S. S., Jani, M. F. M., & Omar, N. A. (2011). An empirical study of success factors of women entrepreneurs in southern region in Malaysia. International Journal of economics and Finance, 3(2), 166-175. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.832.6751&rep=rep1&type=pdf

Al-Mamary, Y. H., & Alshallaqi, M. (2022). Impact of autonomy, innovativeness, risk-taking, proactiveness, and competitive aggressiveness on students' intention to start a new venture. Journal of Innovation & Knowledge, 7(4), 100239. https://doi.org/10.1016/j.jik.2022.100239

Alvarez, S. A., & Barney, J. B. (2007). Discovery and creation: Alternative theories of entrepreneurial action. Strategic entrepreneurship journal, 1(1-2), 11-26. https://doi.org/10.1002/sej.4

Aly, M., Audretsch, D.B. & Grimm, H. Emotional skills for entrepreneurial success: the promise of entrepreneurship education and policy. J Technol Transf 46, 1611–1629 (2021). https://doi.org/10.1007/s10961-021-09866-1

Archer, G., Arndt, F., & Robinson, F. (2022). Value Creation through Environmental Entrepreneurship. In WORLD SCIENTIFIC ENCYCLOPEDIA OF BUSINESS SUSTAINABILITY, ETHICS AND ENTREPRENEURSHIP (pp. 157-173). https://doi.org/10.1142/9789811248863_0006

Ardichvili, A., Cardozo, R., & Ray, S. (2003). A theory of entrepreneurial opportunity identification and development. Journal of Business venturing, 18(1), 105-123. https://doi.org/10.1016/S0883-9026(01)00068-4

Asante, E. A., & Affum-Osei, E. (2019). Entrepreneurship as a career choice: The impact of locus of control on aspiring entrepreneurs' opportunity recognition. Journal of Business Research, 98, 227-235. https://doi.org/10.1016/j.jbusres.2019.02.006

Audretsch, D. B. (2020). Entrepreneurship and culture. Eurasian Economic Review, 10(1), 1-8. https://doi.org/10.1007/s40822-019-00132-2

Baker, H., Kumar, S. & Pandey, N. (2021). A bibliometric analysis of managerial finance: A retrospective. Managerial Finance, 46(11), 1495-1517.

Baker, T., & Nelson, R. E. (2005). Creating something from nothing: Resource construction through entrepreneurial bricolage. Administrative science quarterly, 50(3), 329-366. https://doi.org/10.2189%2Fasqu.2005.50.3.329

Baker, T., & Nelson, R. E. (2005). Creating something from nothing: Resource construction through entrepreneurial bricolage. Administrative science quarterly, 50(3), 329-366. https://doi.org/10.2189%2Fasqu.2005.50.3.329

Bakkar, B. S., Mohammad, A. L., Shayeb, M. A., Suhail, A. Z., & Shendi, Y. A. A. (2021). THE PSYCHOMETRIC CHARACTERISTICS OF THE OMANI ENTREPRENEURIAL SELF-EFFICACY SCALE. Journal of Entrepreneurship Education, 24, 1-12. https://www.proquest.com/openview/7a47f22e198bae0846b8a45d065e7137/1?pq-origsite=gscholar&cbl=28224

Baluku, M. M., Kikooma, J. F., & Kibanja, G. M. (2016). Psychological capital and the startup capital–entrepreneurial success relationship. Journal of Small Business & Entrepreneurship, 28(1), 27-54. https://doi.org/10.1080/08276331.2015.1132512

Barão, A., de Vasconcelos, J. B., Rocha, Á., & Pereira, R. (2017). A knowledge management approach to capture organizational learning networks. International Journal of Information Management, 37(6), 735-740.

Baron, R. A., & Shane, S. (2007). Entrepreneurship: A process perspective. The psychology of entrepreneurship, 19-39. http://dx.doi.org/10.1146/annurev-orgpsych-031413-091326

Barreneche García, A. (2014). Analyzing the determinants of entrepreneurship in European cities. Small Business Economics, 42(1), 77-98. https://doi.org/10.1007/s11187-012-9462-8

Barreneche García, A. (2014). Analyzing the determinants of entrepreneurship in European cities. Small Business Economics, 42(1), 77-98. https://doi.org/10.1007/s11187-012-9462-8

Baum, J. R., & Locke, E. A. (2004). The Relationship of Entrepreneurial Traits, Skill, and Motivation to Subsequent Venture Growth. Journal of Applied Psychology, 89(4), 587–598. https://psycnet.apa.org/doi/10.1037/0021-9010.89.4.587

Bauman, A., & Lucy, C. (2021). Enhancing entrepreneurial education: Developing competencies for success. The International Journal of Management Education, 19(1), 100293. https://doi.org/10.1016/j.ijme.2019.03.005

Bednar, S., Gicheva, D., & Link, A. N. (2021). Innovative activity and gender dynamics. Small Business Economics, 56(4), 1591-1599. https://doi.org/10.1007/s11187-019-00282-2

Begley, T. M., & Boyd, D. P. (1987). Psychological characteristics associated with performence in entrepreneurial firms and smaller businesses. Journal of business venturing, 2(1), 79-93. https://doi.org/10.1016/0883-9026(87)90020-6

Bhaiswar, R., Meenakshi, N., & Chawla, D. (2021). Evolution of Electronic Word of Mouth: A Systematic Literature Review Using Bibliometric Analysis of 20 Years (2000–2020). FIIB Business Review, 10(3), 215-231.

Bhattacharya, S., & Basu, P.K. (1998). Mapping a research area at the micro level using co-word analysis. Scientometrics, 43, 359–372.

Bird, B. (1988). Implementing entrepreneurial ideas: The case for intention. Academy of management Review, 13(3), 442-453. https://doi.org/10.5465/amr.1988.4306970

Birley, S. (1985). The role of networks in the entrepreneurial process. Journal of business venturing, 1(1), 107-117. https://doi.org/10.1016/0883-9026(85)90010-2

Birley, S. (2000). The Role of Networks in the Entrepreneurial. Small Business: Critical Perspectives on Business and Management, 4, 1495. https://books.google.co.in/books?hl=en&lr=&id=8kNx95ta7NoC&oi=fnd&pg=PA1495&dq=an+entrepreneur+is+seeking+the+resources+of+equipment,+space,+money,+advice,+information,+and+reassurance+&ots=rFX0UjTK7t&sig=_T5cLTxmhX9-Jn-3jegOlo2nR9E&redir_esc=y#v=onepage&q=an%20entrepreneur%20is%20seeking%20the%20resources%20of%20equipment%2C%20space%2C%20money%2C%20advice%2C%20information%2C%20and%20reassurance&f=false

Boyack, K.W. & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? Journal of the American Society for Information Science and Technology, 61(12), 2389–2404. https://doi.org/10.1002/asi.21419

Brandstätter, H. (2011). Personality aspects of entrepreneurship: A look at five meta-analyses. Personality and individual differences, 51(3), 222-230. https://doi.org/10.1016/j.paid.2010.07.007

Brown, C., & Thornton, M. (2013). How entrepreneurship theory created economics. Quarterly Journal of Austrian Economics, 16(4). Accessed on 1st Aug,2022: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.430.9684&rep=rep1&type=pdf

Brüderl, J., & Preisendörfer, P. (1998). Network support and the success of newly founded business. Small business economics, 10(3), 213-225. https://doi.org/10.1023/A:1007997102930

Bullough, A., Guelich, U., Manolova, T. S., & Schjoedt, L. (2022). Women's entrepreneurship and culture: gender role expectations and identities, societal culture, and the entrepreneurial environment. Small Business Economics, 58(2), 985-996.

Bullough, A., Renko, M., & Myatt, T. (2014). Danger zone entrepreneurs: The importance of resilience and self–efficacy for entrepreneurial intentions. Entrepreneurship Theory and Practice, 38(3), 473-499. https://doi.org/10.1111%2Fetap.12006

Cacciolatti, L., Lee, S.H. (2015). Entrepreneurial Cognition and Learning. In: Entrepreneurial Marketing for SMEs. Palgrave Macmillan, London. https://doi.org/10.1057/9781137532589_4

Cantamessa, M., Gatteschi, V., Perboli, G., & Rosano, M. (2018). Startups' roads to failure. Sustainability, 10(7), 2346.

Cardon, M. S., Wincent, J., Singh, J., & Drnovsek, M. (2009). The nature and experience of entrepreneurial passion. Academy of management Review, 34(3), 511-532. https://doi.org/10.5465/amr.2009.40633190

Cavallo, A., Ghezzi, A., & Balocco, R. (2019). Entrepreneurial ecosystem research: Present debates and future directions. International Entrepreneurship and Management Journal, 15(4), 1291-1321. https://doi.org/10.1007/s11365-018-0526-3

Chang, Y. C., Chang, H. T., Chi, H. R., Chen, M. H., & Deng, L. L. (2012). How do established firms improve radical innovation performance? The organizational capabilities view. Technovation, 32(7-8), 441-451.

Chen, C. C., Greene, P. G., & Crick, A. (1998). Does entrepreneurial self-efficacy distinguish entrepreneurs from managers?. Journal of business venturing, 13(4), 295-316. https://doi.org/10.1016/S0883-9026(97)00029-3

Coad, A. (2007). Firm growth: a survey. Max Planck Institute of Economics. Available from: http://hal.archivesouvertes.fr/halshs-00155762

Cohen, W.M., & Levinthal, D.A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly, 35(1), 128–152.

Collins, C. J., Hanges, P. J., & Locke, E. A. (2004). The relationship of achievement motivation to entrepreneurial behavior: A meta-analysis. Human performance, 17(1), 95-117. https://doi.org/10.1207/S15327043HUP1701_5

Cope, J. (2005). Toward a dynamic learning perspective of entrepreneurship. Entrepreneurship theory and practice, 29(4), 373-397. https://doi.org/10.1111%2Fj.1540-6520.2005.00090.x

Croitoru, A. (2012). Schumpeter, JA, 1934 (2008), The theory of economic development: An inquiry into profits, capital, credit, interest and the business cycle. Journal of comparative research in anthropology and sociology, 3(02), 137-148.

Dang, H. A. H., & Nguyen, C. V. (2021). Gender inequality during the COVID-19 pandemic: Income, expenditure, savings, and job loss. World Development, 140, 105296.

Davidsson, P. (1991). Continued entrepreneurship: Ability, need, and opportunity as determinants of small firm growth. Journal of business venturing, 6(6), 405-429. https://doi.org/10.1016/0883-9026(91)90028-C

Dej, D. (2010). Defining and measuring entrepreneurial success. Entrepreneurship: A psychological approach, 89-102.

Dimov, D. (2010). Nascent entrepreneurs and venture emergence: Opportunity confidence, human capital, and early planning. Journal of management studies, 47(6), 1123-1153. https://doi.org/10.1111/j.1467-6486.2009.00874.x

Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. Information processing & management, 47(1), 80-96.

Distanont, A., & Khongmalai, O. (2020). The role of innovation in creating a competitive advantage. Kasetsart Journal of Social Sciences, 41(1), 15-21.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285-296.

Doran, J., McCarthy, N., & O'Connor, M. (2018). The role of entrepreneurship in stimulating economic growth in developed and developing countries. Cogent Economics & Finance, 6(1), 1442093. https://doi.org/10.1080/23322039.2018.1442093

Drew, C. H., Pettibone, K. G., Finch, F. O., Giles, D., & Jordan, P. (2016). Automated research impact assessment: A new bibliometrics approach. Scientometrics, 106(3), 987–1005. https://doi.org/10.1007/s11192-015-1828-7

Dubin, R., 1978. Theory Building, second ed. Free Press, New York.

Economic Survey 2021-2022. https://www.ibef.org/economy/economic-survey-2021-22#:~:text=Current%20State%20of%20the%20Economy,8.5%25%20in%202022%2D23.

Eggers, F. (2020). Masters of disasters? Challenges and opportunities for SMEs in times of crisis. Journal of business Research, 116, 199-208.

Eliyana, A., Rohmatul, S., Rohmatul, S., Sridadi, A. R., Razaq, A., & Gunawan, D. R. (2020). The role of motivation on attitudes and entrepreneur achievement. Systematic Reviews in Pharmacy, 11(8), 335-343. https://www.sysrevpharm.org/articles/the-role-of-motivation-on-attitudes-and-entrepreneur-achievement.pdf

Elsafty, A., Abadir, D., & Shaarawy, A. (2020). How does the entrepreneurs' financial, human, social and psychological capitals impact entrepreneur's success. Business and Management Studies, 6(3), 55-71. https://doi.org/10.11114/bms.v6i3.4980

Fabel, O. (2004). Spin-Offs of Entrepreneurial Firms: An O-Ring Approach. Journal of Institutional and Theoretical Economics, 160, 416–438. https://www.jstor.org/stable/40752470

Fayolle, A. (2018). Personal views on the future of entrepreneurship education. In A research agenda for entrepreneurship education. Edward Elgar Publishing.

Fernández-de-las-Peñas, C., Martín-Guerrero, J.D., Florencio, L.L(2023) Clustering analysis reveals different profiles associating long-term post-COVID symptoms, COVID-19 symptoms at hospital admission and previous medical co-morbidities in previously hospitalized COVID-19 survivors. Infection 51, 61–69

Fisher, R., Maritz, A., & Lobo, A. (2014). Evaluating entrepreneurs' perception of success: Development of a measurement scale. International Journal of Entrepreneurial Behavior & Research, 20(5), 478-492. https://doi.org/10.1108/IJEBR-10-2013-0157

Florea, L., Cheung, Y. H., & Herndon, N. C. (2013). For all good reasons: Role of values in organizational sustainability. Journal of business ethics, 114(3), 393-408. https://doi.org/10.1007/s10551-012-1355-x

Ford, S., Garnsey, E., & Probert, D. (2010). Evolving corporate entrepreneurship strategy: technology incubation at Philips. R&d Management, 40(1), 81-90. https://doi.org/10.1111/j.1467-9310.2009.00580.x

Frese, M., & Gielnik, M. M. (2014). The psychology of entrepreneurship. Annual review of organizational psychology and organizational behavior, 1(1), 413-438. https://doi.org/10.1146/annurev-orgpsych-031413-091326

Fried, H. O., & Tauer, L. W. (2015). An entrepreneur performance index. Journal of Productivity Analysis, 44(1), 69-77. https://doi.org/10.1007/s11123-015-0436-0

Gao, Q., Wu, C., Wang, L., & Zhao, X. (2020). The entrepreneur's psychological capital, creative innovation behavior, and enterprise performance. Frontiers in Psychology, 11, 1651.

Gao, S., Meng, F., Gu, Z., Liu, Z., & Farrukh, M. (2021). Mapping and clustering analysis on environmental, social and governance field a bibliometric analysis using Scopus. Sustainability, 13(13), 7304. https://doi.org/10.3390/su13137304

Garg, M., & Gupta, S. (2022). Start-ups and the growing entrepreneurial ecosystem. Journal of Intellectual Property Rights (JIPR), 26(1), 31-38.

Gatewood, E. J., Shaver, K. G., & Gartner, W. B. (1995). A longitudinal study of cognitive factors influencing start-up behaviors and success at venture creation. Journal of business venturing, 10(5), 371-391. https://doi.org/10.1016/0883-9026(95)00035-7

Goel, S., & Karri, R. (2006). Entrepreneurs, Effectual Logic, and Over Trust. Entrepreneurship Theory and Practice, 30 (4), 477-493. https://doi.org/10.1111/j.1540-6520.2006.00131.x

González-Tejero, C.B. and Molina, C.M. (2022), "Training, corporate culture and organizational work models for the development of corporate entrepreneurship in SMEs", Journal of Enterprising Communities: People and Places in the Global Economy,16, 168-188. https://doi.org/10.1108/JEC-12-2021-0178

Gorgievski, M. J., Ascalon, M. E., & Stephan, U. (2011). Small business owners' success criteria, a values approach to personal differences. Journal of Small Business Management, 49(2), 207-232. https://doi.org/10.1111/j.1540-627X.2011.00322.x

Green, S. G., Welsh, M. A., & Dehler, G. E. (2003). Advocacy, performance, and threshold influences on decisions to terminate new product development. Academy of management journal, 46(4), 419-434. https://doi.org/10.5465/30040636

Gupta, N., & Mirchandani, A. (2018). Investigating entrepreneurial success factors of women-owned SMEs in UAE. Management Decision. https://doi.org/10.1108/MD-04-2017-0411

Harzing, A. W., & Alakangas, S. (2016), Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. Scientometrics, 106, 787-804.

Hassan, H., Mahdee, J., Rahman, M. S., & Sade, A. B. (2017). Key success factors and challenges in harnessing the power of women entrepreneurs. World Review of Entrepreneurship, Management and Sustainable Development, 13(4), 450-467.

Hmieleski, K. M., & Baron, R. A. (2009). Entrepreneurs' optimism and new venture performance: A social cognitive perspective. Academy of management Journal, 52(3), 473-488. https://doi.org/10.5465/amj.2009.41330755

Huang, H. C. (2016). Entrepreneurial resources and speed of entrepreneurial success in an emerging market: the moderating effect of entrepreneurship. International Entrepreneurship and Management Journal, 12, 1-26.

Hyder, S. and Lussier, R. (2016) 'Why businesses succeed or fail: A study on small businesses in Pakistan', Journal of Entrepreneurship in Emerging Economics, 8, 82–100.

Ireland, R. D., Hitt, M. A., & Sirmon, D. G. (2003). A model of strategic entrepreneurship: The construct and its dimensions. Journal of management, 29(6), 963-989. https://doi.org/10.1016/S0149-2063(03)00086-2

Iriobe, O. (2018). Entrepreneurial Ventures and Its Components. Entrepreneurship Development: Principles & Practice, Princeton & Associates Publishing Co. Ltd (2018). https://dx.doi.org/10.2139/ssrn.4163901

Jamali, D. (2009), "Constraints and opportunities facing women entrepreneurs in developing countries: a relational perspective", Gender in Management: An International Journal, Vol. 24, 232-251. https://doi.org/10.1108/17542410910961532

Jambulingam, T., Joshi, M. P., & Kathuria, R. (2018). Competitive Analysis of CRM Strategies Using Analytic Hierarchy Process. Management Dynamics, 18(1), 1-22.

Jawabri, A. (2020). The impact of big-5 model leadership traits on team entrepreneurship: An empirical study of small businesses in the UAE. Management Science Letters, 10(3), 497-506. http://dx.doi.org/10.5267/j.msl.2019.9.031

Jayawarna, D., Rouse, J., & Kitching, J. (2013). Entrepreneur motivations and life course. International small business journal, 31(1), 34-56. https://doi.org/10.1177%2F0266242611401444

Jennings, J.E. and Brush, C.G. (2013), "Research on women entrepreneurs: challenges to (and from) the broader entrepreneurship literature?", The Academy of Management Annals, 7 (1), 663-715.

Kaiser, U. (2008). A primer in Entrepreneurship. University of Zurich-Institute for Strategy and Business Economics Ders Notları, 1-17. https://www.business.uzh.ch/dam/jcr:00000000-39c4-f61f-0000-000063375948/Kap5primerentrepreneur2010.pdf

Kakarika, M., Biniari, M., Guillén, L., & Mayo, M. (2022). Where does the heart lie? A multistage process model of entrepreneurial passion and role identity management. Journal of Organizational Behavior. https://doi.org/10.1002/job.2605

Kessler, E. (Ed.) (2013). Encyclopedia of management theory. (Vols. 1-2). SAGE Publications, Ltd., https://doi.org/10.4135/9781452276090

Khan, R. U., Salamzadeh, Y., Shah, S. Z. A., & Hussain, M. (2021). Factors affecting women entrepreneurs' success: a study of small-and medium-sized enterprises in emerging market of Pakistan. Journal of innovation and entrepreneurship, 10(1), 1-21.

Khan, S. H., Majid, A., Yasir, M., & Javed, A. (2021). Social capital and business model innovation in SMEs: do organizational learning capabilities and entrepreneurial orientation really matter?. European Journal of Innovation Management, 24(1), 191-212.

Khan, S., Bhatti, S., Zaman, U., & Hussain, M. (2020). Breaking down the success barrier: The mediating role of absorptive capacity in linking entrepreneurial orientation to IT project success. Pakistan Journal of Commerce and Social Science, 14(2), 529-550.

Kickula, J., Griffiths, M., Bacq, S., & Garud, N. (2018). Catalyzing social innovation: is entrepreneurial bricolage always good?. Entrepreneurship & Regional Development, 30(3-4), 407-420. https://doi.org/10.1080/08985626.2017.1413771

Kiviluoto, N. (2011). Rediscovering profitability in entrepreneurship: Evidence from Finnish high-technology start-ups. Accessed on 1st Aug,2022 http://bibbild.abo.fi/ediss/2011/kiviluoto_niklas.pdf

Kollmann, T., & Christofor, J. (2014). International entrepreneurship in the network economy: Internationalization propensity and the role of entrepreneurial orientation. Journal of international entrepreneurship, 12, 43-66.

Kolstad, I., & Wiig, A. (2015). Education and entrepreneurial success. Small Business Economics, 44(4), 783-796. https://doi.org/10.1007/s11187-014-9621-1

Kremer, M. (1993). The O-Ring theory of economic development. Quarterly Journal of Economics, 108(3), 551–575. https://doi.org/10.2307/2118400

Kurczewska, A., & Mackiewicz, M. (2020). Are jacks-of-all-trades successful entrepreneurs? Revisiting Lazear's theory of entrepreneurship. Baltic Journal of Management, 15(3), 411-430. https://www.emerald.com/insight/1746-5265.htm

Kurniawan, R. Y., Prakoso, A. F., & Soesatyo, Y. (2022). Is Entrepreneurial Intention A Major Factor in The Success of An Entrepreneur?. https://books.google.co.in/books?hl=en&lr=&id=PIOAEAAAQBAJ&oi=fnd&pg=PA242&dq=an+entrepreneur%E2%80%99s+behavior+impacts+many+factors,+such+as+readiness+for+innovation,+proactive+personality,+generalized+self-efficacy,+stress+tolerance,+need+for+autonomy,+and+locus+of+control,+which+have+a+significant+correlation+with+business+creation+an&ots=X4bLIFOJFv&sig=r8dzbSc1ugeqPobsgXm9Q4QBlQQ&redir_esc=y#v=onepage&q&f=false

Kyrö, P. (2018). The conceptual contribution of education to research on entrepreneurship education. In A research agenda for entrepreneurship education (pp. 164-186). Edward Elgar Publishing.

Lazear, E. P. (2004). Balanced skills and entrepreneurship. The American Economic Review, 94(2), 208–211. https://doi.org/10.1257/0002828041301425

Lévi-Strauss, C. (1962). La pensée sauvage. https://monoskop.org/images/f/f6/Levi-Strauss_Claude_La_Pensee_sauvage_1962.pdf

Li, Y., & Ahlstrom, D. (2020). Risk-taking in entrepreneurial decision-making: A dynamic model of venture decision. Asia Pacific Journal of Management, 37(3), 899-933. https://doi.org/10.1007/s10490-018-9631-7

Liguori, E., & Winkler, C. (2020). From offline to online: Challenges and opportunities for entrepreneurship education following the COVID-19 pandemic. Entrepreneurship Education and Pedagogy, 3(4), 346-351.

Liu, M., Gorgievski, M. J., Qi, J., & Paas, F. (2022). Increasing teaching effectiveness in entrepreneurship education: Course characteristics and student needs differences. Learning and Individual Differences, 96, 102147.

Lounsbury, M., Cornelissen, J., Granqvist, N., & Grodal, S. (2019). Culture, innovation and entrepreneurship. Innovation, 21(1), 1-12.

Lu, Y., Zhou, L., Bruton, G., & Li, W. (2010). Capabilities as a mediator linking resources and the international performance of entrepreneurial firms in an emerging economy. Journal of International Business Studies, 41(3), 419-436.

Lumpkin, G. T., & Dess, G. G. (1996). Clarifying the entrepreneurial orientation construct and linking it to performance. Academy of management Review, 21(1), 135-172. https://doi.org/10.5465/amr.1996.9602161568

Lumpkin, G. T., & Dess, G. G. (2001). Linking two dimensions of entrepreneurial orientation to firm performance: The moderating role of environment and industry life cycle. Journal of business venturing, 16(5), 429-451. https://doi.org/10.1016/S0883-9026(00)00048-3

Luthans, F., Youssef, C. M., & Avolio, B. J. (2007). Psychological capital: Investing and developing positive organizational behaviour. Positive organizational behaviour, 1(2), 9-24.

Malecki, E. J. (2018). Entrepreneurs, networks, and economic development: A review of recent research. Reflections and extensions on key papers of the first twenty-five years of advances. https://doi.org/10.1108/S1074-754020180000020010

Maslakçı, A., Sürücü, L., & Şeşen, H. (2022). Positive psychological capital and university students' entrepreneurial intentions: does gender make a difference?. International Journal for Educational and Vocational Guidance, 1-26.

Maslakçı, A., Sürücü, L., & Şeşen, H. (2022). Relationship between positive psychological capital and entrepreneurial intentions of university students: the mediating role of entrepreneurial self-efficacy. World J Entrep Manag Sustain Dev, 18, 305-319.

Mazzei, M. J. (2018). Strategic entrepreneurship: Content, process, context, and outcomes. International Entrepreneurship and Management Journal, 14(3), 657-670. https://doi.org/10.1007/s11365-018-0512-9

Mazzucato, M. (2011). The Entrepreneurial State:''The state has not just fixed markets, but actively created them…". Demos: London. http://dx.doi.org/10.1177/1024529416678084

McGee, J. E. M., & Peterson, S. L. Mueller, & JM Sequeira. 2009. Entrepreneurial Self-Efficacy: Refining the measure. Entrepreneurship Theory and Practice, 33(4). https://doi.org/10.1111%2Fj.1540-6520.2009.00304.x

McKenzie, D., & Paffhausen, A. L. (2019). Small firm death in developing countries. Review of economics and statistics, 101(4), 645-657. https://doi.org/10.1162/rest_a_00798

Minniti, M. and Naudé, W. (2010), "What do we know about the patterns and determinants of female entrepreneurship across countries?", The European Journal of Development Research, 22 (3) 277-293.

Muis, I., Suleeman, J., & Riyanti, B. P. D. (2017, September). " Why do women have less opportunity for entrepreneurial success compared to men?"(Case study: Indonesian entrepreneurship). In 2nd International Conference on Education, Science, and Technology (ICEST 2017) (pp. 49-51). Atlantis Press. https://dx.doi.org/10.2991/icest-17.2017.17

Mumu, J. R., Saona, P., Haque, M. S., & Azad, M. A. K. (2021). Gender diversity in corporate governance: a bibliometric analysis and research agenda. Gender in Management: An International Journal. https://doi.org/10.1108/GM-02-2021-0029

Nanda, R. (2016). Financing high-potential entrepreneurship. IZA World of Labor. https://doi.org/10.15185/izawol.252

Naudé, W. (2013). Entrepreneurship and economic development: Theory, evidence and policy. Evidence and Policy. IZA Discussion Paper, (7507). https://dx.doi.org/10.2139/ssrn.2314802

Ndofirepi, T. M. (2020). Relationship between entrepreneurship education and entrepreneurial goal intentions: psychological traits as mediators. Journal of Innovation and Entrepreneurship, 9(1), 1-20. https://doi.org/10.1186/s13731-020-0115-x

Newman, A., Ucbasaran, D., Zhu, F. E. I., & Hirst, G. (2014). Psychological capital: A review and synthesis. Journal of organizational behavior, 35(S1), S120-S138. https://doi.org/10.1002/job.1916

Ojo, O. D., Petrescu, M., Petrescu, A. G., & Bîlcan, F. R. (2017). Impact of innovation on the entrepreneurial success: Evidence from Nigeria. African Journal of Business Management, 11(12), 261-265. https://doi.org/10.5897/AJBM2017.8277

Olafsen, E., & Cook, P. A. (2016). Growth entrepreneurship in developing countries: A preliminary literature review. https://documents1.worldbank.org/curated/zh/587331562049723102/pdf/Growth-Entrepreneurship-in-Developing-Countries-A-Preliminary-Literature-Review.pdf

Olivas, D., Mukhopadhyay, S., & Frankwick, G. L. (2022). Effects of ethnicity and gender on ethically driven small business performance. Journal of Small Business and Enterprise Development, (ahead-of-print). https://doi.org/10.1108/JSBED-09-2021-0383

Orlandi, L. B. (2017). Am I an entrepreneur? Identity struggle in the contemporary women entrepreneurship discourse. Contemporary Economics, 11(4), 487-499. https://www.econstor.eu/bitstream/10419/195507/1/102921915X.pdf

Orser, D & Dyke, L. (2009, June 22). The influence of gender and occupational-role on entrepreneurs'and corporate managers' success criteria. The Free Library. [Html- Dokument]. Available under: http://www.thefreelibrary.com

Palich, L. E., & Bagby, D. R. (1995). Using cognitive theory to explain entrepreneurial risk-taking: Challenging conventional wisdom. Journal of business venturing, 10(6), 425-438. https://doi.org/10.1016/0883-9026(95)00082-J

Panda, S. (2018). Constraints faced by women entrepreneurs in developing countries: review and ranking. Gender in Management: An International Journal. https://doi.org/10.1108/GM-01-2017-0003

Parker, S. C. (2018). The economics of entrepreneurship. Cambridge University Press. https://doi.org/10.1017/9781316756706

Pathirana, K. P. J. M., & Semasinghe, D. M. (2018). Entrepreneurial traits and venture growth in Sri Lanka. http://192.248.73.38/bitstream/handle/94ousl/1685/Vistas%20Dec-18_edited01082019-103-126.pdf?sequence=1

Perez, E. H., & Canino, R. M. B. (2009). The importance of the entrepreneur's perception of "success". Review of International Comparative Management, 10(5), 990-1010. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.861.8978&rep=rep1&type=pdf

Politis, D. (2005). The process of entrepreneurial learning: A conceptual framework. Entrepreneurship theory and practice, 29(4), 399-424. https://doi.org/10.1111%2Fj.1540-6520.2005.00091.x

Qian, J., Law, R., & Wei, J. (2019). Knowledge mapping in travel website studies: A scientometric review. Scandinavian Journal of Hospitality and Tourism, 19(2), 192-209. https://doi.org/10.1080/15022250.2018.1526113

Ramos-Rodriguez, A. R., Medina-Garrido, J. A., Lorenzo-Gómez, J. D., & Ruiz-Navarro, J. (2010). What you know or who you know? The role of intellectual and social capital in opportunity recognition. International small business journal, 28(6), 566-582.

Ramos-Rodrıguez, Antonio-Rafael, Jose-Aurelio Medina-Garrido, Jose-Daniel Lorenzo-Gomez, and Jose Ruiz-Navarro. 2010. "What You Know or Who You Know? The Role of Intellectual and Social Capital in Opportunity Recognition." International Small Business Journal 28 (6): 566582. doi:10.1177/0266242610369753.

Rastogi, M., Baral, R. and Banu, J. (2022), "What does it take to be a woman entrepreneur? Explorations from India", Industrial and Commercial Training, 54, 333-356. https://doi.org/10.1108/ICT-03-2021-0022

Rauch, A., & Frese, M. (2007). Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success. European Journal of work and organizational psychology, 16(4), 353-385. https://doi.org/10.1080/13594320701595438

Rauch, A., Wiklund, J., Lumpkin, G. T., & Frese, M. (2009). Entrepreneurial orientation and business performance: Cumulative empirical evidence. https://doi.org/10.1111/j.1540-6520.2009.00308.x

Raza, S., Minai, M. S., Abrar ul Haq, M., & Zain, A. Y. M. (2018). Entrepreneurial network towards small firm performance through dynamic capabilities: the conceptual perspective. Academy of Entrepreneurship Journal, 24(4), 1-9.

Razmus, W., & Laguna, M. (2018). Dimensions of entrepreneurial success: A multilevel study on stakeholders of micro-enterprises. Frontiers in psychology, 9, 791. https://doi.org/10.3389/fpsyg.2018.00791

Red Global Entrepreneurship Monitor (GEM). Informe Gem España 2018–2019. https://www.cise.es/pdf/informe-gem-2018-19.pdf (accessed on 1 April 2021).

Rosa, P., Carter, S., & Hamilton, D. (1996). Gender as a determinant of small business performance: Insights from a British study. Small business economics, 8(6), 463-478. https://doi.org/10.1007/BF00390031

Rosca, E., Agarwal, N., & Brem, A. (2020). Women entrepreneurs as agents of change: A comparative analysis of social entrepreneurship processes in emerging markets. Technological Forecasting and Social Change, 157, 120067. https://doi.org/10.1016/j.techfore.2020.120067

Rylková, Ž., & Čemerková, Š. (2020). Personality Assumptions and Attitudes Towards Business. In Developing Entrepreneurial Competencies for Start-Ups and Small Business (pp. 59-72). IGI Global. https://doi.org/10.4018/978-1-7998-2714-6.ch004

Saes, AM, & Marcovitch, J. (2020). Entrepreneurial education: recent trajectory and challenges. Journal of Entrepreneurship and Small Business Management, 9 (1), 1-9.

Sajjad, M., Kaleem, N., Chani, M. I., & Ahmed, M. (2020). Worldwide role of women entrepreneurs in economic development. Asia Pacific Journal of Innovation and Entrepreneurship, 14(2), 151-160.

Salisu, I., Hashim, N., Mashi, M. S., & Aliyu, H. G. (2020). Perseverance of effort and consistency of interest for entrepreneurial career success: Does resilience matter?. Journal of Entrepreneurship in Emerging Economies, 12(2), 279-304.

Saptono, A., & Najah, S. (2018). Development of an assessment instrument of affective domain for entrepreneurship in senior high school. Journal of Entrepreneurship Education, 21(4), 1-12. https://www.proquest.com/openview/ea316265d08a847b645834ae05c9dc5d/1?pq-origsite=gscholar&cbl=28224 .

Sarasvathy, S. D., Simon, H. A., & Lave, L. (1998). Perceiving and managing business risks: Differences between entrepreneurs and bankers. Journal of Economic Behavior and Organization, 33, 207–225. https://doi.org/10.1016/S0167-2681(97)00092-9

Sarasvathy, S.D. (2001). Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. Academy of Management Review, 26, 243–263. https://doi.org/10.5465/amr.2001.4378020

Sardana, V. & Singhania, S (2022). Fifty Years of Research in Deposit Insurance: A Bibliometric Analysis and Review, FIIB Business Review, Volume ahead of Print, Issue ahead of print.

Schumpeter, J. A. (1934). The Theory of Economic Development Cambridge Mass. First published in German in 1911SchumpeterThe theory of economic development1934. http://compaso.eu/wpd/wp-content/uploads/2013/01/Compaso2012-32-Croitoru.pdf

Schumpeter, J., & Backhaus, U. (2003). The theory of economic development (pp. 61-116). Springer US.

Senyard, J., Baker, T., & Davidsson, P. (2009). Entrepreneurial bricolage: Towards systematic empirical testing. Frontiers of Entrepreneurship Research, 29(5), 5. http://digitalknowledge.babson.edu/fer/vol29/iss5/5

Shaikh, E., Tunio, M. N., & Qureshi, F. (2022). Finance and women's entrepreneurship in DETEs: A literature review. Entrepreneurial Finance, Innovation and Development, 191-209.

Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. Academy of management review, 25(1), 217-226. https://doi.org/10.5465/amr.2000.2791611

Sharma, G. D., Kraus, S., Liguori, E., Bamel, U. K., & Chopra, R. (2022). Entrepreneurial Challenges of COVID-19: Re-thinking entrepreneurship after the crisis. Journal of Small Business Management, 1-23.

Shaver, K. G., & Scott, L. R. (1992). Person, process, choice: The psychology of new venture creation. Entrepreneurship theory and practice, 16(2), 23-46. https://doi.org/10.1177%2F104225879201600204

Shepherd, D. A., & DeTienne, D. R. (2005). Prior knowledge, potential financial reward, and opportunity identification. Entrepreneurship theory and practice, 29(1), 91-112. https://doi.org/10.1111%2Fj.1540-6520.2005.00071.x

Simon, A., Bartle, C., Stockport, G., Smith, B., Klobas, J. E., & Sohal, A. (2015). Business leaders' views on the importance of strategic and dynamic capabilities for successful financial and non-financial business performance. International Journal of Productivity and Performance Management, 64(7), 908-931.

Singh, A. K., Kota, H. B., Sardana, V., & Singhania, S. (2021b). Does gender diversity on board promote corporate social responsibility? An empirical analysis of sustainable development goals. Australasian Accounting, Business and Finance Journal, 15(5), 22-40.

Singh, A. K., Singhania, S., & Sardana, V. (2019). Do women on boards affect firm's financial performance? Evidence from Indian IPO firms. Australasian Accounting, Business and Finance Journal, 13(2), 53-68.

Singh, J., Singhania, S., & Aggrawal, D. (2021a). Gender diversity on corporate boards: Review and future research agenda through bibliometric mapping. Corporate Governance & Sustainability Review, 5(3), 57-72.

Singh, J., Singhania, S., & Aggrawal, D. (2023). Does board gender diversity impact financial performance? Evidence from the Indian IT sector. Society and Business Review, 18(1), 51-70.

Singhania, S., Singh, J. and Aggrawal, D., (2022a). Board committees and financial performance: Exploring the effects of gender diversity in the emerging economy of India. International Journal of Emerging Markets, Volume ahead of Print, Issue ahead of print.

Singhania, S., Singh, J., & Aggrawal, D. (2023a). Gender diversity on board and corporate sustainability: a quantitative review based on bibliometric mapping. International Journal of System Assurance Engineering and Management, 14(1), 267-286.

Singhania, S., Singh, J., Aggrawal, D., & Rana, S. (2023b). Board gender diversity and sustainability reporting quality: a generalized ordered logit approach. Kybernetes, Volume ahead of Print, Issue ahead of print.

Singhania, S., Singh, R. K., Singh, A. K., & Sardana, V. (2022a). Corporate governance and risk management: A bibliometric mapping for future research agenda. Indian Journal of Corporate Governance, 15(2), 223-255.

Siu, W. S., & Lo, E. S. (2013). Cultural contingency in the cognitive model of entrepreneurial intention. Entrepreneurship Theory & Practice, 37(2), 147–173.

Slavec, A., & Prodan, I. (2012). The influence of entrepreneur's characteristics on small manufacturing firm debt financing. Journal for East European Management Studies, 104-130. https://www.jstor.org/stable/23281745

Somsuk, N., & Laosirihongthong, T. (2014). A fuzzy AHP to prioritize enabling factors for strategic management of university business incubators: Resource-based view. Technological forecasting and social change, 85, 198-210. https://doi.org/10.1016/j.techfore.2013.08.007

Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., & Schoder, D. (2016). Business model development, founders' social capital and the success of early stage internet start-ups: a mixed-method study. Information Systems Journal, 26(5), 421-449.

Stam, W., & Elfring, T. (2008). Entrepreneurial orientation and new venture performance: The moderating role of intra-and extraindustry social capital. Academy of management journal, 51(1), 97-111. https://doi.org/10.5465/amj.2008.30744031

Staniewski, M. W., & Awruk, K. (2019). Entrepreneurial success and achievement motivation–A preliminary report on a validation study of the questionnaire of entrepreneurial success. Journal of Business Research, 101, 433-440. https://doi.org/10.1016/j.jbusres.2019.01.073

Staniewski, M. W., & Awruk, K. (2021). Parental attitudes and entrepreneurial success. Journal of Business Research, 123, 538-546. https://doi.org/10.1016/j.jbusres.2020.10.039

Stel, A. V., Carree, M., & Thurik, R. (2005). The effect of entrepreneurial activity on national economic growth. Small business economics, 24, 311-321.

Stephan, U. (2022). Cross-cultural innovation and entrepreneurship. Annual Review of Organizational Psychology and Organizational Behavior, 9, 277-308.

Strielkowski, W., Krejcí, M., & Čabelková, I. (2015). Factors that influence the success of small and medium enterprises in ICT: a case study from the Czech Republic. Business: Theory and Practice, 16(3), 304-315.

Tang, J. J. (2020). Psychological capital and entrepreneurship sustainability. Frontiers in Psychology, 11, 866.

Toms, S., Wilson, N., & Wright, M. (2020). Innovation, intermediation, and the nature of entrepreneurship: A historical perspective. Strategic Entrepreneurship Journal, 14(1), 105-121. https://doi.org/10.1002/sej.1310

Trevelyan, R. (2008). Optimism, overconfidence and entrepreneurial activity. Management Decision, 46(7), 986-1001. https://doi.org/10.1108/00251740810890177

Tur-Porcar, A., Roig-Tierno, N., & Llorca Mestre, A. (2018). Factors affecting entrepreneurship and business sustainability. Sustainability, 10(2), 452. https://doi.org/10.3390/su10020452

Uddin, M. K., & Akther, S. (2019). Contributions of Social Entrepreneurs: A Study from the Perspective of the Society, Culture, and Environment of Bangladesh. Management Dynamics, 19(1), 1-18.

Vaiz, T., & Ekemen, M. A. (2022). The Role of Positive Psychological Capital and Personality Traits on the Relationship of Conscious Awareness and Entrepreneurial Intention. Revista de Cercetare si Interventie Sociala, 78.

Van Praag, C. M., & Versloot, P. H. (2007). What is the value of entrepreneurship? A review of recent research. Small business economics, 29(4), 351-382. https://doi.org/10.1007/s11187-007-9074-x

Van-Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538.

Vieira, E.S. & Gomes, J.A. (2009). A comparison of Scopus and Web of Science for a typical university. Scientometrics, 81, 587.

Wach, D., Stephan, U., & Gorgievski, M. (2016). More than money: Developing an integrative multi-factorial measure of entrepreneurial success. International Small Business Journal, 34(8), 1098-1121. https://doi.org/10.1177%2F0266242615608469

Wagner, J. (2003). Testing Lazear's jack-of-all-trades view of entrepreneurship with German micro data. Applied Economics Letters, 10(11), 687-689. https://doi.org/10.1080/1350485032000133273

Wales, W. J., Covin, J. G., & Monsen, E. (2020). Entrepreneurial orientation: The necessity of a multilevel conceptualization. Strategic Entrepreneurship Journal, 14(4), 639-660.

Wang, C., Mundorf, N., & Salzarulo-McGuigan, A. (2022). Entrepreneurship education enhances entrepreneurial creativity: The mediating role of entrepreneurial inspiration. The International Journal of Management Education, 20(2), 100570.

Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. Research Policy, 46(8), 1416–1436. https://doi.org/10.1016/j.respol.2017.06.006

Watson, J. (2012). Networking: Gender differences and the association with firm performance. International Small Business Journal, 30(5), 536-558. https://doi.org/10.1177%2F0266242610384888

Woodfield, P. J., Shepherd, D., & Woods, C. (2017). How can family winegrowing businesses be sustained across generations? International Journal of Wine Business Research, 29(2), 122-139. https://doi.org/10.1108/IJWBR-12-2015-0052

World Economic Forum (2013) https://reports.weforum.org/new-models-for-entrepreneurship/executive-summary.

Xie, G. H., Wang, L. P., & Lee, B. F. (2021). Understanding the impact of social capital on entrepreneurship performance: the moderation effects of opportunity recognition and operational competency. Frontiers in Psychology, 12, 687205.

Yin, W., Moon, H. C., & Lee, Y. W. (2019). The success factors of Korean global start-ups in the digital sectors through internationalization. International Journal of Global Business and Competitiveness, 14(1), 42-53.

Yusof, S. W. M., Jabar, J., Murad, M. A., & Ortega, R. T. (2017). Exploring the cultural determinants of entrepreneurial success: The case of Malaysia. International journal of advanced and applied sciences, 4(12), 287-297. https://doi.org/10.21833/ijaas.2017.012.048

Zhao, H., O'Connor, G., Wu, J., & Lumpkin, G. T. (2021). Age and entrepreneurial career success: A review and a meta-analysis. Journal of Business Venturing, 36(1), 106007.

Zhou, M., Zhou, Y., Zhang, J., Obschonka, M., & Silbereisen, R. K. (2019). Person–city personality fit and entrepreneurial success: An explorative study in China. International journal of psychology, 54(2), 155-163. https://doi.org/10.1002/ijop.12451

Zou, Huan, Xiaoyun Chen, Long Wai Rico Lam, and Xiaohui Liu. 2015. “Psychological Capital and Conflict Management in the Entrepreneur Venture Capitalist Relationship in China: The Entrepreneur Perspective.” International Small Business Journal. Online publication ahead of print; 122. doi:10.1177/0266242614563418.

Zupic, I. and Cater, T. (2015). Bibliometric methods in management and organization. Organisational Research Methods, 18, 429–472.

The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-023-12093-7

ORIGINAL ARTICLE

Modeling and simulation of material removal characteristics in magnetorheological shear thickening polishing

Zhen Ma[1] · Yebing Tian[1,2] · Cheng Qian[1] · Shadab Ahmad[1,3] · Zenghua Fan[1,2] · Zhiguang Sun[1]

Received: 23 December 2022 / Accepted: 31 July 2023
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

Abstract
Magnetorheological shear thickening polishing (MRSTP) was experimentally demonstrated to achieve nano-level surface quality in previous studies. The aim of work is to further investigate numerical simulation and material removal characteristics during the MRSTP process. The apparatus of magnetic field generation was designed and developed. The model of magnetic field distribution was imported to computational fluid dynamics (CFD) module to generate desired polishing environment. Polishing stress behaviors were analyzed based on various polishing parameters through the CFD simulation results. It was found that the pressure on the workpiece surface decreased with the increase of the rotational speed of the workpiece while it increased with the increase of the flow velocity of the MRSTP media. A continuous increment in shear stress on the workpiece surface was observed with the increase of the rotation speed of the workpiece and the flow velocity of the MRSTP media. Based on the characteristics of polishing pressure, material removal rate (MRR) model of MRSTP process was established. A serial of MRSTP experiments were conducted on aluminum 6061. Surface roughness (Sa) of 79.0 nm was obtained, which was improved over 77%. The simulated model and MRR model were experimentally validated. The average error between theoretical and experimental results was 4.1%. It can provide a useful guidance for the novel MRSTP process.

Keywords Magnetorheological shear thickening polishing · Numerical simulation · Force characteristics · Material removal model

1 Introduction

Magnetic field assisted polishing (MFAP) methods have become one of research hotspots due to their flexibility, self-sharpening and self-adaptability for the recent years. MFAP methods are widely employed for polishing various hard-to-process materials, such as silicon, sapphire, glass, ceramic, superalloy. Barman et al. [1] investigated the influence of magnetorheological (MR) polishing fluids with different compositions on polishing performance. Two different MR fluids were utilized to polish the bio-titanium alloy to obtain hydrophilic and hydrophobic surfaces, respectively. Zhai et al. [2] proposed ultrasound assisted magnetorheological polishing (UAMP) on sapphire and established a material removal rate (MRR) model of UAMP based on the Preston equation. The feasibility of this model was experimentally verified. Luo et al. [3] conducted polishing experiments on zirconia ceramics using MRP method. Surface roughness R_a of zirconia ceramics decreased to 0.702 nm from initial 71.976 nm using diamond abrasive particles. The maximum material removal rate of 1 mg/min was attained. Yamaguchi et al. [4] employed magnetic abrasive finishing (MAF) method to improve the tool life of uncoated WC/Co cutting tool. It was concluded that the cutting tool life was improved by 150%. Li et al. [5] performed a hybrid post-processing of MAF and heat treatment (HT) on the surface of 3D-printed Inconel 718 alloy. The surface roughness Ra of the workpiece decreased to 0.15 μm from initial 2.7 μm. The hybrid post-treatment of MAF and HT demonstrated that the mechanical properties such as yield strength and elongation of the Inconel 718 alloy manufactured in additive form

✉ Yebing Tian
tianyb@sdut.edu.cn; tyb79@sina.com

[1] School of Mechanical Engineering, Shandong University of Technology, 266 Xincun West Road, Zibo, Shandong 255049, People's Republic of China

[2] Institute for Advanced Manufacturing, Shandong University of Technology, Zibo 255049, People's Republic of China

[3] Centre for Advanced Production and Industrial Engineering Research, Delhi Technological University, Delhi 110042, India

Published online: 05 August 2023

were significantly improved. Umehara et al. [6] designed and fabricated a novel magnetic floating polishing platform for large-size and large-volume silicon nitride (Si_3N_4) balls. The polishing process was divided into three stages i.e., roughing stage, intermediate stage and finishing stage. The surface roughness Ra of the polished silicon nitride (Si_3N_4) balls decreased to 6.70 nm from initial 937.14 nm.

For the MFAP method, a lot of numerical simulations were employed to understand the effects of processing parameters on the polishing performance before the experiments were conducted. This facilitated the simplification of tests and efficiency improvement. Kumar et al. [7] investigated the mechanism of magnetorheological polishing on miniature tooth contours where the forces on the surface of the micro-gear were simulated using COMSOL software. A surface roughness model of polished micro-gear was established using simulation results. Wang et al. [8] proposed a new magnetic field assisted mass polishing technology for efficient polishing of multiple free-form parts simultaneously. The effect of magnetic field on the material removal characteristics was analyzed by the finite element method. The factors affecting surface generation were investigated by polishing experiments. The results showed that magnetic field assisted mass polishing could effectively polish multiple free-form parts with nano-level surface finish. Liu et al. [9] simulated the three-dimensional profile of the polishing points during magnetorheological polishing. A material removal rate (MRR) model considering pressure and shear stress was developed using the simulation results, which was verified for accuracy. Sharma [10] simulated the magnetic field strength of the newly designed magnetorheological polishing equipment. A mathematical model of the relationship between surface roughness and magnetic field strength was established. The results showed that the surface roughness decreased with the increase of magnetic field strength. Grover et al. [11] designed a magnetorheological polishing equipment for polishing cylindrical inner surfaces using radially curved permanent magnets. Simulation of the magnetic field strength of this radially curved permanent magnet was conducted. A mathematical model for predicting surface roughness was developed based on the results of the simulation. After validation, the maximum error of the model was 8.06%. Guo et al. [12] designed a novel magnetorheological polishing equipment with Halbach array for quartz specimens. The magnetic field strength of Halbach array was simulated to verify its large area and high strength. The surface roughness of the quartz specimen eventually improved up to 0.544 nm after polishing.

An accurate material removal model is of benefit to provide theoretical guidance and reveal the polishing mechanism. Great efforts have been made to establish the predictive material removal model for the MFAP. Zhai et al. [2] developed a material removal model for magnetorheological polishing where the coupling effect of magnetic and vibrational fields was considered. It was verified that the relative error between the theoretical and experimental results was less than 20%. Chen et al. [13] established a two-dimensional material removal model for magnetorheological polishing based on hydrodynamic theory and the Preston equation. The rationality of the material removal model was verified by comparing the experimental results with the simulation results. The error between the model and the experimental results was 9.8%. Ming et al. [14] established a material removal model for a non-Newtonian fluid polishing method based on weak magnetorheological strengthening thickening effects. The model was developed using non-Newtonian fluid kinematics. The average error between the model and experimental results was 4.45%. Li et al. [15] established a material removal model for the magnetic abrasive finishing method using fluid abrasives in which the effect of magnetic fields on material removal was considered. This model agrees to the experimental results with an average error of 4.51%.

In our previous studies, magnetorheological shear thickening polishing (MRSTP) was proposed as a novel magnetic field assisted polishing process which combined the magnetorheological effect with the shear thickening effect together [16–19]. As one of loosed abrasive polishing methods, MRSTP possesses excellent adaptability and processing efficiency. It has a magnetorheological effect that enables the rapid formation of magnetically enhanced particle clusters in the polishing media under the excitation of a magnetic field [16]. These clusters exert much stronger holding force on the abrasive particles, which makes the polishing process more efficient than the traditional MFAP. Additionally, MRSTP has a shear thickening effect during polishing, which results in larger shear stress on the workpiece surface than the conventional MRF. Due to its high adaptability and efficiency, MRSTP has been used to polish various difficult-to-machine materials [16–19]. MRSTP media was prepared by mixing a certain proportion of polyethylene glycol (PEG 200), SiO_2, abrasives, carbonyl iron powder (CIP) and chemical additives [16]. Simulation of the magnetic field strength of the magnetic field generation device was performed to verify the rationality of the structure [17]. The magnetic field generator and the modified magnetic abrasive were used to polish the titanium alloy Ti-6Al-4 V. Surface roughness of workpiece reduced from 1.17 μm to 54 nm. Also, in a case MRSTP media were specially developed for titanium alloy under the action of a four-pole rotating magnetic field [18]. The surface roughness of titanium alloy Ti-6Al-4 V was reduced from 169 to 61 nm and the effectiveness of this polishing media was verified. A new magnetic field generation device was designed which was capable of generating alternating magnetic fields for the additive manufacturing of 361L rectangular micro-fabricated structures[19]. Material removal rate models were developed and processing experiments

were conducted. The results showed that the top and bottom of the rectangular microstructure were uniformly removed from the material. The feasibility of machining additively manufactured workpieces with MRSTP media was verified.

Nevertheless, recent studies on MRSTP were focused on processing experiments and simulation of magnetic fields. The coupling simulation of magnetic and flow fields in the MRSTP process with CFD methods was not explored. Therefore, in this work, CFD simulations of the designed MRSTP apparatus were carried out using Maxwell and Fluent modules in ANSYS software. It aimed to investigate the pattern of influence of polishing parameters on the polishing force applied to the workpiece surface. A mathematical model between MRR and polishing parameters was developed using the output of the normal forces from the CFD simulation results. The accuracy of the MRR model and the validity of the simulation results were verified by experiments.

2 Principles

Figure 1 shows the principle of MRSTP. MRSTP media are distributed inside polishing bowl which is fixed on the rotary table of either CNC machine or industrial robot. The magnetic field generation device is mounted below the polishing bowl. The workpiece is clamped onto the spindle bottom of either CNC machine or industrial robot. Under the action of the magnetic field, the MRSTP media form magnetic-force-reinforced flexible abrasive clusters. As the magnetic field has adsorption effect on the CIP, the shear rate of the MRSTP media will decrease when they flow through the magnetic field. This is macroscopically expressed as an increase in the apparent viscosity of the MRSTP media. The MRSTP media move in a circular motion driven by the rotary polishing bowl. The CIP and abrasive grains in the magnetic-force-reinforced flexible abrasive clusters are constantly renewed and pass through the immersed workpiece. The relative velocity between the MRSTP media and the workpiece is easily controlled by the spindle and the rotary table of either CNC machine or industrial robot.

3 Design of magnetic field generation device

Figure 2 shows the three-dimensional model and two-dimensional model of the magnetic field generation device. The device comprises of two magnetic pole mounts and twelve magnetic poles. The two pole mounts were made of Steel 45 with twelve holes (i.e., each mount had six poles) for holding the magnetic poles. The six poles were divided into three sets i.e., the top set, the middle set and the bottom set. The middle set were designed at an angle of 120° to the bottom sets. The material of all twelve poles was chosen as N38. The distribution of the magnetic field was simulated using Maxwell software. The materials defined in the model for the magnetic field generation device were same as those used in the actual polishing process. Three lines were planned for the measurement of magnetic field strength as shown in Fig. 2. The magnetic field intensity along the design path was measured using a Gauss meter (Model: GM 500).

Figure 3 shows the results of the magnetic field simulation. It can be seen that the magnetic force lines uniformly distributed in the polished area while magnetic field intensity were still strong around the workpiece. The maximum magnetic field strengths of 525.7 mT and 527.1 mT were obtained at the starting points of lines 1 and 3, respectively. Along line 1 and line 3, the magnetic field strength decreased with the increase of the distance. A maximum magnetic field strength of 183.3 mT was obtained at 3 mm away from the starting point of the line 2. As shown in Fig. 3 (a), since the magnetic force lines at the starting point of line 2 was not close to the magnetic field generating device, the magnetic field strength along line 2 firstly increased and then decreased with the distance increasing from the starting point.

Figure 4 shows the comparisons of magnetic field intensity between the simulation and experimental results. It

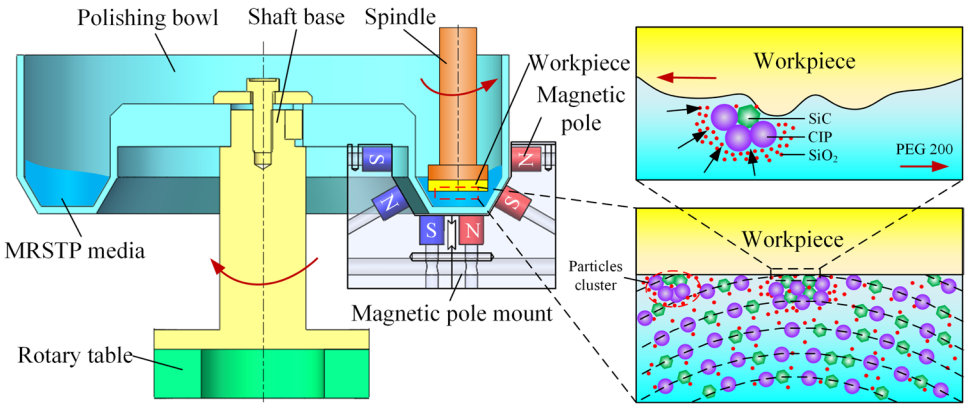


Fig. 1 Principle of the MRSTP

can be seen that the measured magnetic field strengths are consistent with the simulation values. The measured values at the starting point of the line 1 and line 3 were 315.4 mT and 367.0 mT, respectively. The measured value was 125.3 mT at 3 mm away from the starting point of the line 2. The validity of the simulation was verified.

4 Numerical simulation of flow field

4.1 Computational fluid dynamics methods

The flow field of the MRSTP process was numerically simulated using Ansys Fluent. Mixture models including STFs, CIP, SiC

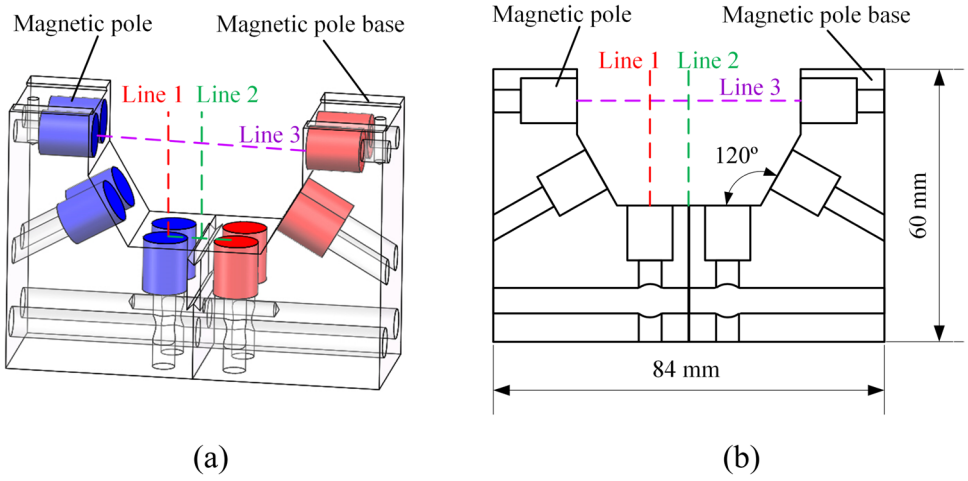


Fig. 2 Model of magnetic field generating device. (**a**) 3D model, (**b**) 2D model

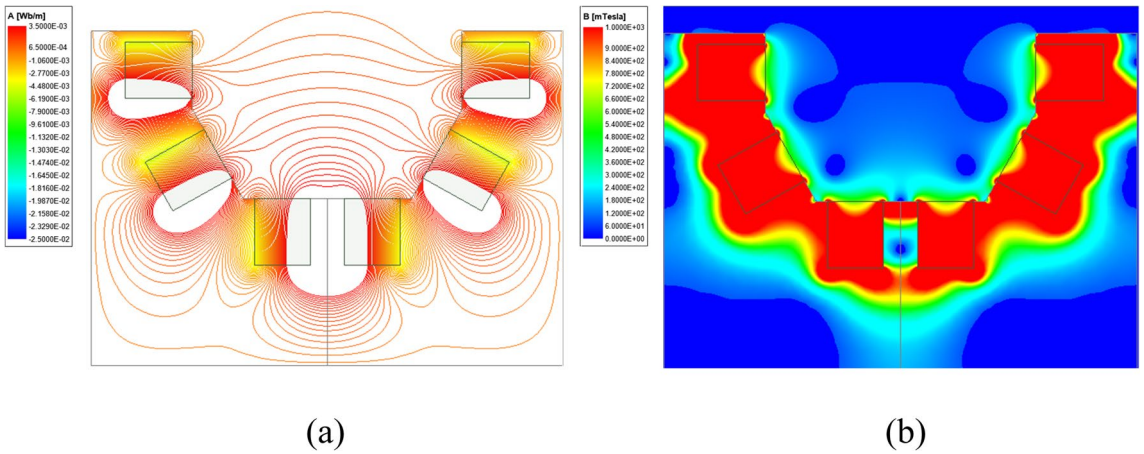


Fig. 3 Maxwell simulation results. (**a**) Distribution of magnetic force lines, (**b**) Distribution of magnetic field intensity

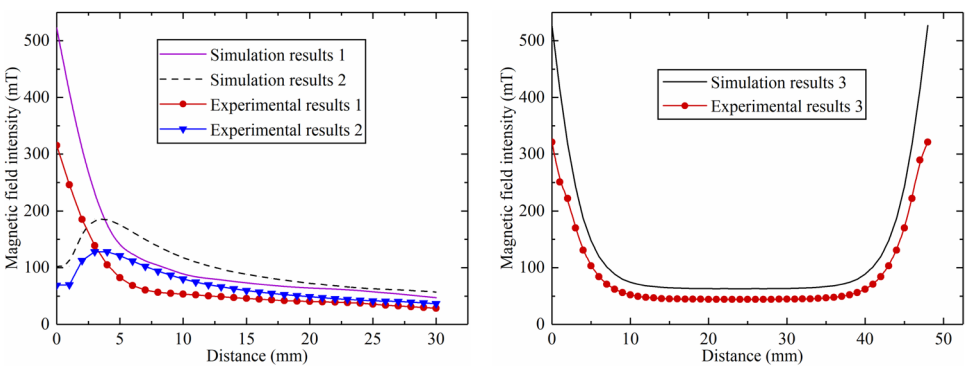


Fig. 4 Comparisons of magnetic field intensity between simulations and experiments

and other multiphase blends were used. The output from Maxwell was coupled using the MHD module in Fluent to simulate the magnetic field environment in the flow field. The following assumptions were made during this numerical simulation: (i) The workpiece was an ideal rigid material. (ii) The processing environment temperature was controlled at 25℃. In the MRSTP process, it was assumed that the fluid was incompressible and operated in a constant temperature environment. Hence, the continuity and momentum equations for incompressible viscous fluids were used to describe the motion state of the MRSTP media [20]:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \tag{1}$$

$$\begin{cases}\frac{\partial u}{\partial t}+(V\cdot\nabla)u=f_x-\frac{1}{\rho}\frac{\partial p}{\partial x}+\frac{1}{\rho}\left(\frac{\partial \sigma_{xx}}{\partial x}+\frac{\partial \tau_{yx}}{\partial y}+\frac{\partial \tau_{zx}}{\partial z}\right)\\ \frac{\partial v}{\partial t}+(V\cdot\nabla)v=f_y-\frac{1}{\rho}\frac{\partial p}{\partial y}+\frac{1}{\rho}\left(\frac{\partial \sigma_{yy}}{\partial y}+\frac{\partial \tau_{zy}}{\partial z}+\frac{\partial \tau_{xy}}{\partial x}\right)\\ \frac{\partial w}{\partial t}+(V\cdot\nabla)w=f_z-\frac{1}{\rho}\frac{\partial p}{\partial z}+\frac{1}{\rho}\left(\frac{\partial \sigma_{zz}}{\partial z}+\frac{\partial \tau_{xz}}{\partial x}+\frac{\partial \tau_{yz}}{\partial y}\right)\end{cases} \tag{2}$$

where V is the velocity of the fluid. u, v, and w are the components of the fluid velocity in the x, y, and z directions. ρ is the density of the fluid. p is the pressure of the fluid. f is the mass force acting on the fluid. σ and τ are stress components.

The magnetic force on the fluid in the MHD model can be calculated by the following equation:

$$\begin{cases}\vec{f}_{m,x}=\mu_0\left(\vec{M}_x\frac{\partial H_x}{\partial x}+\vec{M}_y\frac{\partial H_x}{\partial y}+\vec{M}_z\frac{\partial H_x}{\partial z}\right)\\ \vec{f}_{m,y}=\mu_0\left(\vec{M}_x\frac{\partial H_y}{\partial x}+\vec{M}_y\frac{\partial H_y}{\partial y}+\vec{M}_z\frac{\partial H_y}{\partial z}\right)\\ \vec{f}_{m,z}=\mu_0\left(\vec{M}_x\frac{\partial H_z}{\partial x}+\vec{M}_y\frac{\partial H_z}{\partial y}+\vec{M}_z\frac{\partial H_z}{\partial z}\right)\end{cases} \tag{3}$$

where μ_0 is the vacuum permeability. M_x, M_y and M_z are the components of the magnetization intensity in the x, y, and z directions. H_x, H_y and H_z are the component of the magnetic field intensity in the x, y, and z directions.

The motion pattern of the fluid can be obtained from the calculation formula of Reynolds number Re:

$$Re=\frac{\rho\bullet v\bullet d_{re}}{\mu} \tag{4}$$

where ρ is the density of the fluid, v is the velocity of the fluid, μ is the dynamic viscosity of the fluid, d_{re} is the characteristic length of the fluid. The flow state of the MRSTP media was determined to be laminar based on the experimental parameters calculated as $\mathrm{Re}<2320$. The SIMPLE algorithm was used to solve the coupled velocity–pressure problem. Since the diffusion effect in the MRSTP media flow process could be neglected, it was a pure convection problem and the discrete format for establishing the discrete equations adopted the second order windward format.

4.2 Geometric model for numerical simulation

Figure 5 shows the geometric model for the numerical simulation in the software of Ansys Fluent. In order to reduce the computational effort and facilitate the computational convergence, the circular polishing bowl of MRSTP was designed as a linear polishing trough. This linear polishing trough had a velocity inlet on one side and a pressure outlet on the other side. The wall of the polishing trough was defined as non-slip wall surface. The wall of the workpiece was set up as rotational wall surface. As the top surface of the polishing trough was MRSTP medium, it was defined as symmetric boundary. In order to improve the computational accuracy, the polished trough was unstructured meshed. The grid size was chosen as 1 mm while the grid was encrypted near the workpiece with size of 0.5 mm. The magnetic field simulation results from the Sect. 3 were further processed and loaded into the software of Ansys Fluent. The loading results are shown in Fig. 5 (c). There was not any distortion observed in the magnetic field and the loading results were accurate. Two lines, AB and CD, were set up on the workpiece surface as the output locations of pressure and shear force in the simulation results, respectively.

4.3 Input parameters for numerical simulation

The simulation input parameters are listed in Table 1. The effects of with the rotational speed of workpiece and the flow velocity of the MRSTP media on the normal pressure and shear force on the workpiece surface were investigated. The maximum magnetic field strength was obtained along line 2 at 3 mm from the starting point as shown in Fig. 4 (a). Considering the existence of excitation gap and the thickness of polishing bowl in the actual processing, the gap between the workpiece and the bottom of the polishing trough was set up as 1 mm.

The rheological profiles of the STFs were measured in order to give shear thickening effect to the MRSTP media during the numerical simulation [16]. When the shear rate is less than the critical shear rate, STFs show shear thinning action. When the shear rate is between the critical shear rate and the maximum shear rate, the STFs show shear thickening action [21–23]. When the shear rate is greater than the maximum shear rate, the STFs show a shear thinning action. The viscosity of the STFs was defined in FLUENT using this rheological curve via programming with C language (Fig. 6).

4.4 Analysis of simulation results

4.4.1 Variation of normal pressure

In the Preston equation, the material removal rate during polishing is associated to the normal pressure on the surface of the workpiece and the velocity of the relative motion

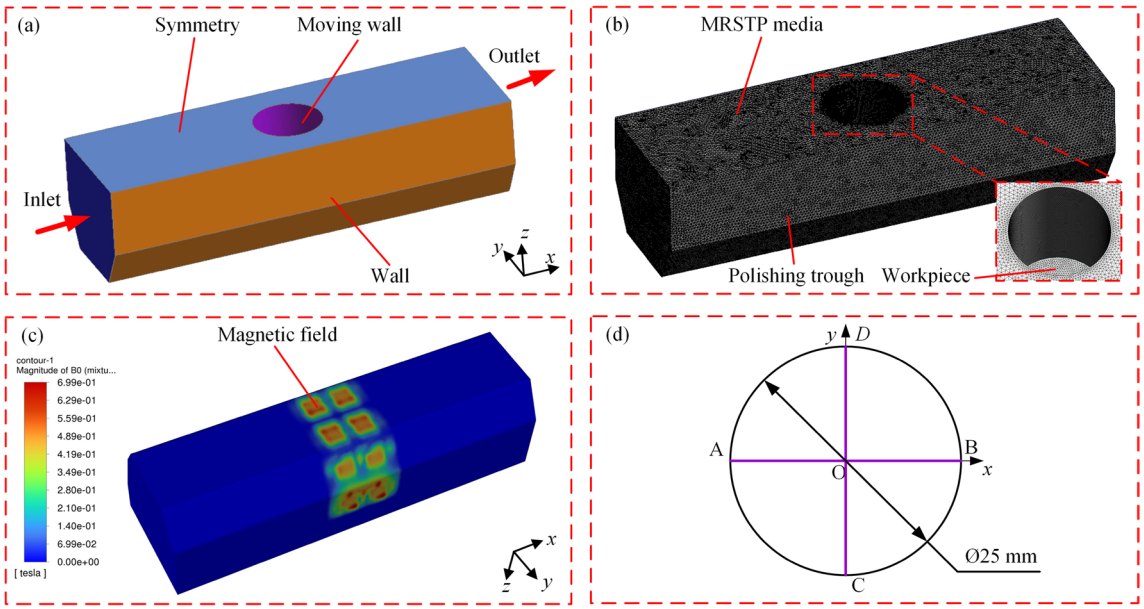


Fig. 5 Simulation models. (**a**) 3D model of a linear polishing trough, (**b**) Grid division, (**c**) Results of magnetic field loading, (**d**) Position of the force measurement on the surface of the workpiece

Table 1 Input parameters used for simulation study

Items	Values
Workpiece rotational speed (rpm)	1000, 1500, 2000
Flow velocity of MRSTP media (m/s)	0.8, 1.2, 1.6,
Working gap (mm)	1
Density of STFs (kg/m^3)	1796.64
Volume fraction of CIP	5.1%
Volume fraction of SiC	4.2%
Workpiece material	6061

between the workpiece and the polishing fluid [24]. Therefore, it is essential to investigate the normal pressure distribution and variation on the workpiece surface through numerical simulation of the flow field.

Figure 7 shows the variation of normal pressure distribution versus different workpiece rotational speeds with the increase of working distance at the MRSTP media velocity of 0.8 m/s. It can be seen that the normal pressure distribution on the surface of the workpiece decreased along the *x*-direction and almost unchanged along the *y*-direction. This main reason was because the MRSTP media flowed along the *x*-direction. When the MRSTP media flowed under the workpiece, the height of the flow channel suddenly became narrow that made the MRSTP media to gather over one side of the workpiece and generated great pressure on the workpiece surface. Meanwhile, due to the high kinetic energy of the MRSTP media, the strong impact produced a large normal pressure on the surface of the workpiece when the

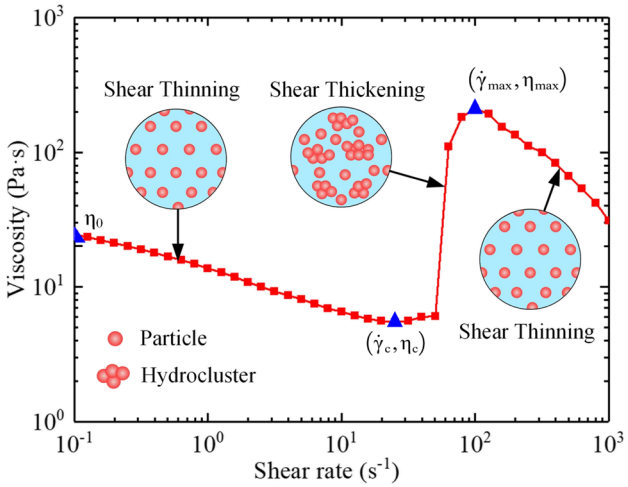


Fig. 6 Variation of STFs viscosity with shear rate

media got to touch the workpiece surface. As the kinetic energy of the MRSTP media was gradually consumed with the flow of MRSTP media over the workpiece surface, the normal pressure on the workpiece surface slightly decreased.

The normal pressure on the workpiece surface gradually decreased as the workpiece rotational speeds increased from 1000 to 2000 rpm. As the rotation speeds of the workpiece increase, the MRSTP media underneath the workpiece will be driven by the workpiece to increase the flow rate. In this case, the normal pressure on the surface of the workpiece decreased, which was in accordance with the description of Bernoulli's principle.

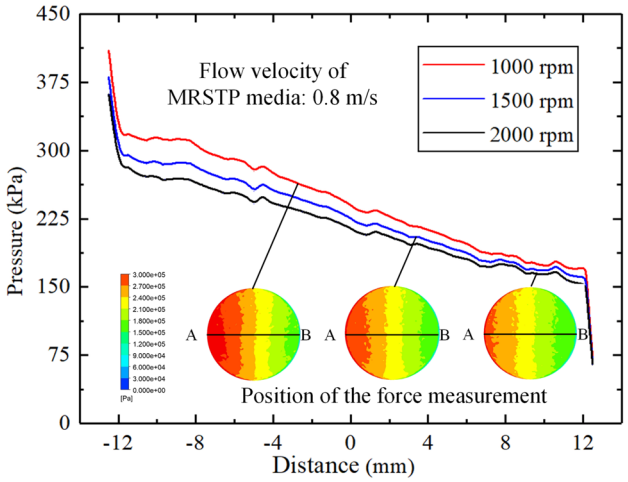


Fig. 7 Variation of normal pressure distribution versus workpiece rotational speed

Figure 8 shows the variation of normal pressure distribution versus the MRSTP media flow velocity of 0.8 m/s, 1.2 m/s and 1.6 m/s with the increase of working distance at the workpiece rotational speed of 2000 rpm. With the increase of flow velocity of MRSTP media, the normal pressure on the surface of the workpiece gradually increased as the kinetic energy became large.

4.4.2 Variation of shear stress

Shear stress is mainly caused by the fluid viscosity when the fluid media flow through the workpiece contact surface. In the ultraprecision polishing with non-newton fluid, material removal from the workpiece surface is not only related to the normal pressure but also the shear stress on the workpiece.

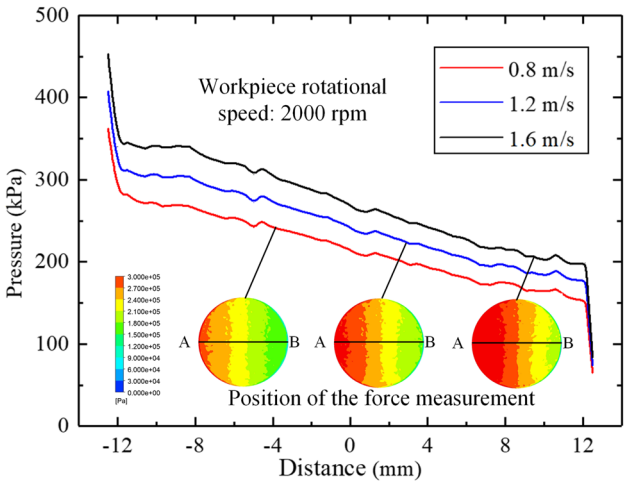


Fig. 8 Variation of normal pressure distribution versus flow velocity of MRSTP media

Therefore, it is also important to investigate the effect of processing parameters on the shear stress.

Figure 9 shows the variation of shear stress distribution on the workpiece surface along the *y*-direction versus different workpiece rotational speed with the increase of working distance the MRSTP media flow velocity of 0.8 m/s. It was seen that the minimum shear stress value appeared near the low side of the workpiece. This was attributed to the combined effects between workpiece rotation and MRSTP media flow. During the numerical simulation, the flow direction of the MRSTP media was set along *x*-direction and the workpiece rotational speed was defined as counter clockwise as labelled in Fig. 9. Hence, along the *y*-axis, the linear velocity of the workpiece and the flow velocity of the MRSTP media gradually varied from the same direction to the opposite direction. This caused that the minimum shear rate in the polishing process occurred close to the point C. When the workpiece rotational speed increased, the shear stress near the D point increased. Meanwhile, the location of the minimum shear stress on the workpiece surface gradually moved upward.

Figure 10 shows the variation of shear stress distribution versus MRSTP media flow velocity of 0.8 m/s, 1.2 m/s, and 1.6 m/s with the increase of working distance at the workpiece rotational speed of 2000 rpm. As the MRSTP media flow velocity increased, the shear stress near the D point increased. The location of the minimum shear stress on the workpiece surface gradually shifted downward.

It was found that the maximum normal pressure on the workpiece surface occurred at the point A while the maximum shear stress occurred at the point D in the MRSTP as shown in Figs. 9 and 10. With the increase the workpiece rotational speed, the normal pressure on the workpiece surface decreased while the shear stress increased. It might lead to nonuniform polishing due to the upward shift of

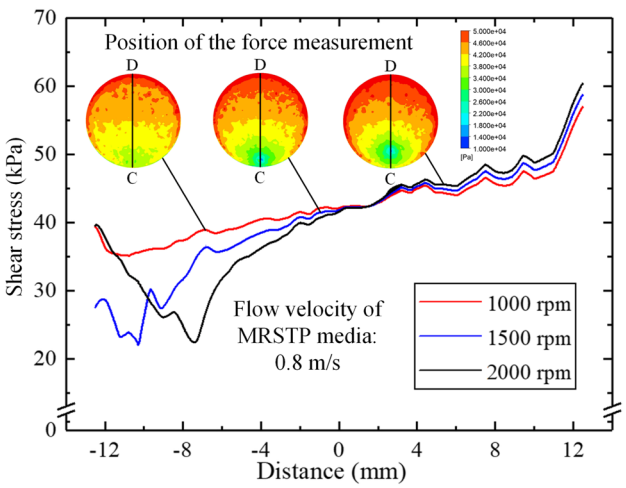


Fig. 9 Variation of shear stress versus workpiece rotational speed

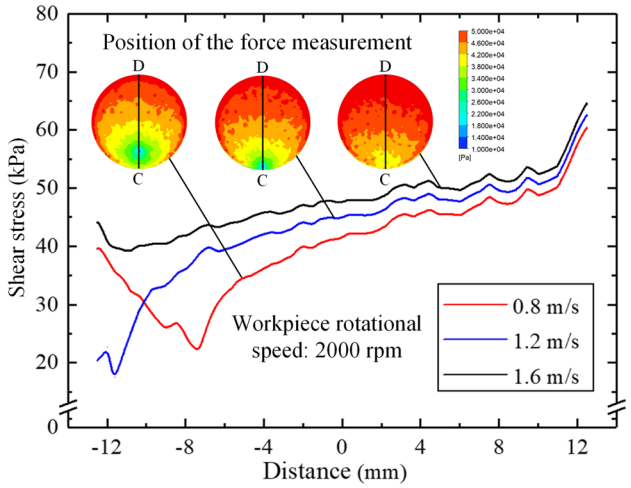


Fig. 10 Variation of shear stress versus flow velocity of MRSTP media

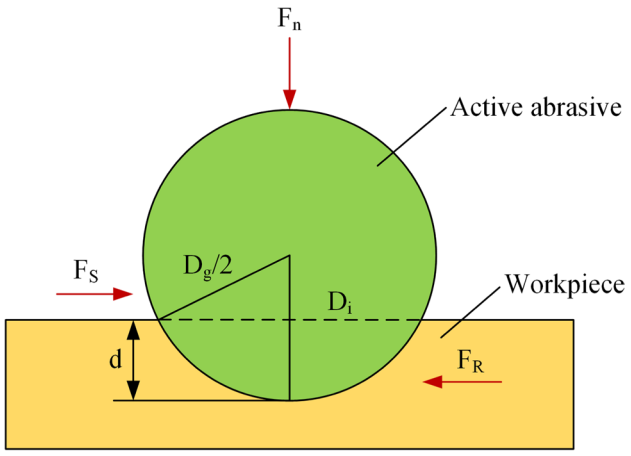


Fig. 11 Illustration of acting force of a single grain on workpiece surface

the minimum shear stress location. As the MRSTP media flow velocity increased, both the normal pressure and shear forces on the workpiece surface became large. Due to the downward shift of the minimum shear stress location, the polishing uniformity of the workpiece surface was positively affected. Hence, in the actual MRSTP process, the polishing efficiency and uniformity would be improved with low workpiece rotational speed and high the MRSTP media flow velocity.

5 Modeling of material removal rate

5.1 Forces in the MRSTP process

Figure 11 shows illustration of the acting force of a single abrasive grain on workpiece surface during MRSTP process. Under the action of the normal pressure F_n, the abrasive grain generates an indentation depth d on the workpiece surface. Because of the relative movement between the workpiece and the media fluid, the abrasive grain is subjected to the shear force F_s. When the reacting force F_R is less than the shear force F_s, the relative motion of the abrasive grain occurs to achieve the material removal on the workpiece [25, 26].

According to the Preston equation [27–29] as expressed as below.

$$MRR = k \cdot p \cdot v \tag{5}$$

where k is the Preston coefficient, p is the normal pressure applied to the workpiece surface, and v is the instantaneous relative velocity of the workpiece surface at the point of contact with the media, the material removal rate is a function of the normal pressure on the workpiece surface and the relative velocity between the workpiece and the media. The normal pressure on the workpiece surface during MRSTP consist of the three components i.e. the fluid dynamic pressure P_d, the floating force of the polishing media P_f and the magnetic pressure P_m [14]:

$$p = p_d + p_f + p_m \tag{6}$$

As both the rheological properties of the STFs and the magnetic field on the kinematic state of the MRSTP media were considered in the numerical simulation (Sect. 4. 2), the pressure in the numerical simulation was the sum of the mentioned the three different pressures viz: P_d, P_f and P_m.

According to the simulation results, the normal pressure distribution on the workpiece surface was almost unchanged along the y-axis direction while the normal pressure on the workpiece surface decreased along the x-axis direction. To simplify the calculation, the surface of the workpiece was divided into N bar areas as shown in Fig. 12.

When N is infinite, all these N regions could be considered as rectangular regions. At this point, the normal pressure in the region D_i ($i = 1, 2, 3 \ldots N$) are equal at each point, the normal pressure was defined as p_i ($i = 1, 2, 3 \ldots N$). Therefore, the total pressure on the region D_i can be calculated as:

$$p_{\mathrm{i}} = \overline{p}_i \cdot S_{Di} \tag{7}$$

$$S_{Di} = \frac{50}{N} \cdot \sqrt{25^2 - \left(25 - \frac{25}{N} i\right)^2} \tag{8}$$

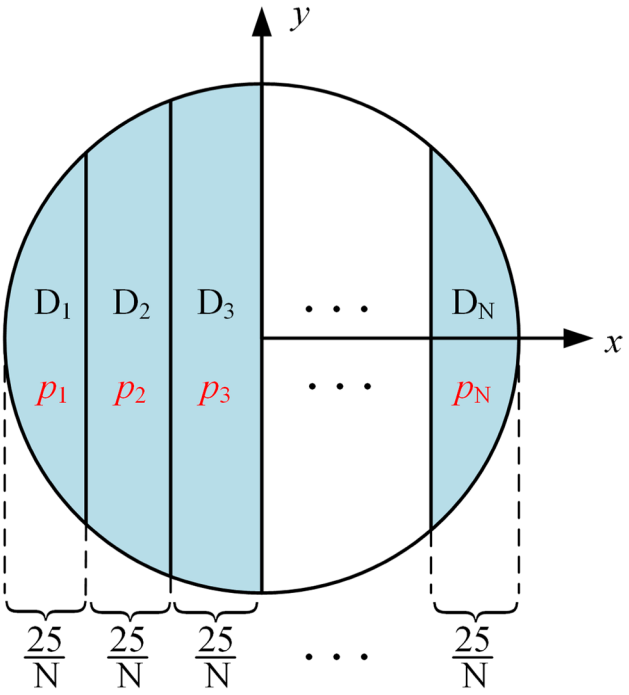


Fig. 12 Area division of the workpiece surface

where S_{Di} is the area of region D_i. Thus, when N tended to infinity, the total pressure on the workpiece surface can be expressed as:

$$p = \lim_{N\to+\infty} \sum_{i=1}^{N} p_i \tag{9}$$

Substituting Eq. (5), (6) into Eq. (7), it was yielded as:

$$p = \lim_{N\to+\infty} \sum_{i=1}^{N} \overline{p}_i \cdot \frac{1250}{N} \cdot \sqrt{\frac{i}{N}\left(2-\frac{i}{N}\right)} \tag{10}$$

5.2 Relative velocity in the MRSTP process

Figure 13 shows a simplified kinematic model of the MRSTP. O_1 is the rotation center of the spindle and O_2 is the rotation center of the rotary table. The xO_1y coordinate system is established with O_1 as the coordinate origin. ρ is the radius of the polishing bowl and ρ_0 is the eccentric distance between the spindle and the rotary table. It is assumed that the workpiece rotated around the spindle with angular velocity ω_1 and the polishing bowl rotated around the rotary table with angular velocity ω_2. The linear velocity of any point P on the surface of the workpiece is $\boldsymbol{v_1}$ about O_1 and $\boldsymbol{v_2}$ about O_2. Hence, the sum of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ is the instantaneous relative velocity $\boldsymbol{v}$ at the point of contact between the

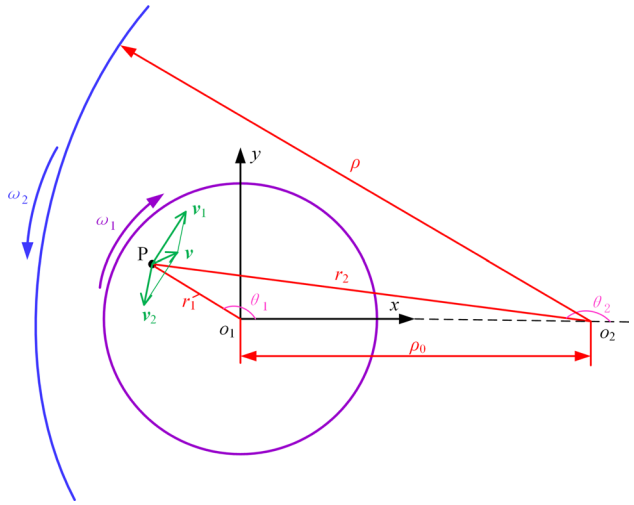


Fig. 13 A simplified kinematic model of the MRSTP

workpiece surface and the MRSTP media. The bit vectors of point P about O_1 and O_2 are defined as $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, respectively. The included angles of the line between P and O_1 and the line between P and O_2 in the positive direction of the x-axis were θ_1 and θ_2, respectively.

The relative velocity of the workpiece and MRSTP media at point P can be expressed as:

$$\vec{v} = \vec{v}_1 + \vec{v}_2 \tag{11}$$

$$\begin{cases} \vec{v}_1 = -\vec{r}_1 \cdot \omega_1 \\ \vec{v}_2 = \vec{r}_2 \cdot \omega_2 \end{cases} \tag{12}$$

The velocities $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ along the x and y axes are decomposed, respectively. Meanwhile, according to the formula for relative velocity v and the geometric relationship as shown in Fig. 13, the formulas are deduced as below.

$$v = \sqrt{v_x^2 + v_y^2} \tag{13}$$

$$r_2 = \sqrt{\left(r_1 \sin\theta_1\right)^2 + \left(\rho_0 - r_1\cos\theta_1\right)^2} \tag{14}$$

$$\cos\left(\theta_1 - \theta_2\right) = \frac{r_1^2 + r_2^2 - \rho_0^2}{2r_1 r_2} \tag{15}$$

The instantaneous relative velocity of the contact point between the workpiece surface and the MRSTP media is obtained as below [30]:

$$v = \sqrt{r_1^2\left(\omega_1 - \omega_2\right)^2 + \rho_0^2\omega_2^2 + 2r_1\rho_0\left(\omega_1 - \omega_2\right)\omega_2\cos\theta_1} \tag{16}$$

Fig. 14 MRSTP experimental setup

5.3 Material removal rate model in MRSTP process.

Based on the Preston equation, the theoretical material removal rate for a planar workpiece surface is expressed as:

$$MRR_t = \iint_D k \cdot p \cdot v d\sigma \tag{17}$$

Substituting Eq. (8) and Eq. (14) into Eq. (15), the Eq. (16) is yielded as below.

$$MRR_t = k \lim_{N\to+\infty} \sum_{i=1}^{N} \overline{p}_i \cdot \frac{2 \cdot 25^2}{N} \cdot \sqrt{\frac{i}{N}\left(2-\frac{i}{N}\right)} \cdot \iint_D \sqrt{r_1^2(\omega_1-\omega_2)^2 + \rho_0^2\omega_2^2 + 2r_1\rho_0(\omega_1-\omega_2)\omega_2\cos\theta_1} d\sigma \tag{18}$$

where k is the Preston coefficient, which is a constant determined by the processing conditions including magnetic field. N is the number of workpiece surface division areas. I ($i=1, 2, 3\ldots N$) is the serial number of the area. p_i is the pressure at any point on the area i, obtained from the simulation results. r_1 is the distance from any point P on the workpiece to the center of rotation of the spindle. ω_1 and ω_2 are the rotational speeds of the spindle and rotary table, respectively. ρ_0 is the eccentric distance between the spindle and the rotary table. θ_1 is the angle between the line connecting the point P and the center of rotation of the spindle with the positive direction of the x-axis. Equation (16) showed that the MRR is related to the MRSTP geometry configuration and process variables. The key process variables are the rotational speeds of the spindle and rotary table and working gaps.

Table 2 Experimental parameters

Items	Parameters
Spindle rotational speed (rpm)	1000, 1250, 1500, 1750, 2000
Rotational speed of rotary table (rpm)	25, 50, 75, 100, 125
Working gap (mm)	1
CIP particle size (μm)	18
SiC particle size (μm)	13
Weight ratio (CIP: SiC)	3:1
STFs concentration (wt %)	20

6 Experimental details

To verify the accuracy of the material removal model and the validity of the simulation results, a series of experiments was carried out. The experimental setup was built up on a CNC machining center (VKN640), as

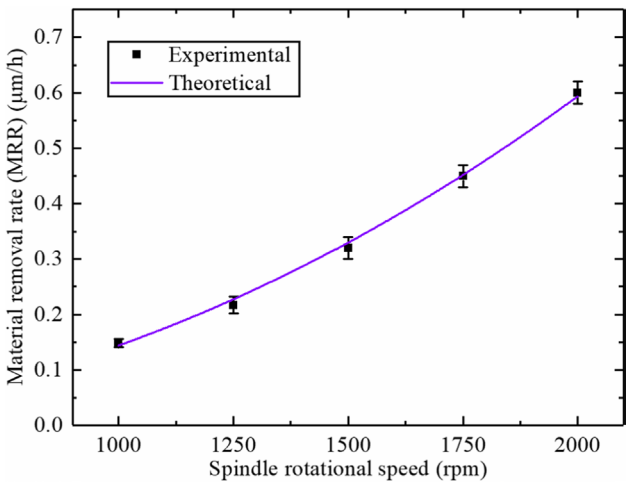


Fig. 15 Influence of spindle rotational speed on MRR

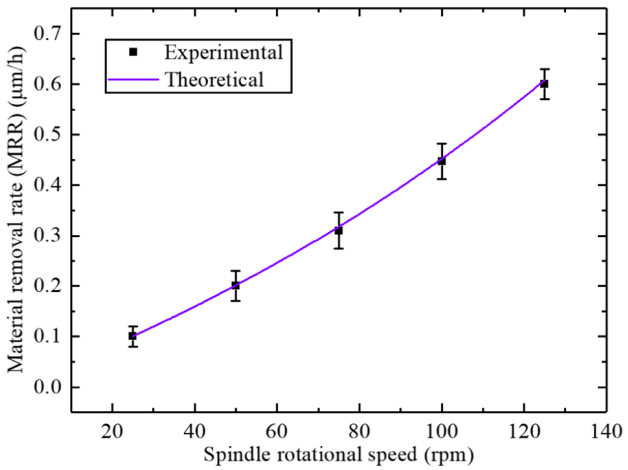


Fig. 16 Influence of rotational speeds of rotary table on MRR

shown in Fig. 14. The magnetic field generation device was installed on a vertical slipway to adjust the working gap. The vertical slipway was held to the CNC worktable by a magnetic base. The workpiece was fixed on the spindle. The polishing bowl was installed on the rotary table. The MRSTP media were contained in the polishing bowl. Magnetic-enhanced particle clusters were formed under the action of magnetic field. The workpiece relative motion was generated with the magnetic-enhanced particle cluster to achieve material removal via the rotation of the workpiece and bowl [31].

A circular Aluminum alloy 6061 with diameter of 25 mm and thickness of 5 mm was used as the workpiece. The initial (S_a) value of the workpiece surface was 337.63 nm. The working gap between the workpiece and the polishing bowl was chosen as 1.0 mm. The concentration of STF base fluid was 20 wt% in the MRSTF media. For comparison with the simulation results, the flow velocity of the MRSTP media in the simulation parameters were converted into the rotational speed of the rotary table. The detailed experimental parameters and their values are shown in Table 2.

7 Results and discussion

7.1 Verification of material removal rate model

The theoretical material removal rate (MRR_t) of MRSTP was compared with the experimental material removal rate (MRR_e). MRR_t was calculated by Eq. (16). MRR_e was calculated by Eq. (17) as below:

$$MRR_e = \frac{M_i - M_p}{S \cdot \rho \cdot t} \tag{19}$$

where, M_i was the initial mass of the workpiece, M_p was the mass of the workpiece after MRSTP, S was the area of the polished plane of the workpiece, ρ was the density of the workpiece, t was the polishing time of MRSTP.

The error Er between MRR_t and MRR_e could be calculated by Eq. (18) [32]:

$$Er = \left| \frac{MRR_t - MRR_e}{MRR_e} \right| \times 100\% \tag{20}$$

The effect of different spindle rotational speeds on MRR was investigated, as shown in Fig. 15. As the spindle rotational speed increased, the MRR increased approximately as a power function. This phenomenon was consistent with the expression of the Preston equation. As the relative velocity of the workpiece and the media increased with the increase of spindle rotational speed, the MRR became large. Although the workpiece surface pressure decreased as the spindle rotational speed increased (see Fig. 7), the reduction in normal pressure played less role in the MRR than the relative velocity.

Figure 16 shows the effects of rotational speeds of the rotary table on the MRR. It was seen that the MRR also increased with the increase of rotational speed of the rotary table. According to Eq. (15), the relative velocity of the workpiece and MRSTP media became large with the increase of rotational speeds of rotary table. Meanwhile, according to Fig. 8, the normal workpiece pressure also increased as rotational speed of the rotary table increased. This resulted in large MRR.

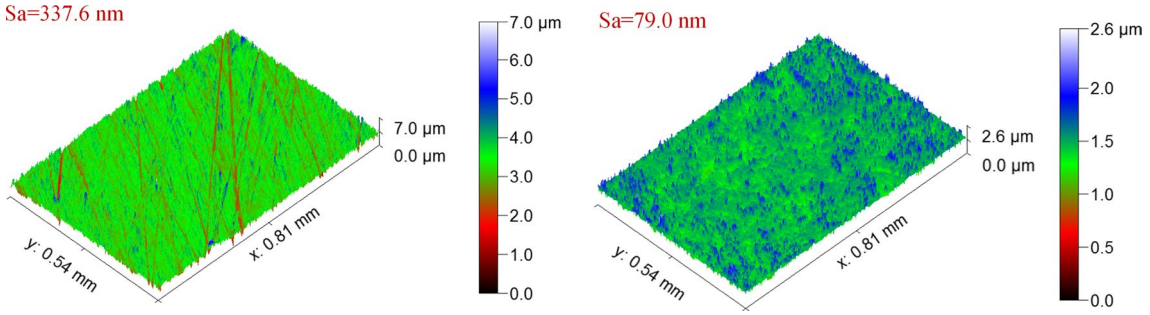


Fig. 17 3D topography of workpiece surface before and after MRSTP

According to Fig. 15 and Fig. 16, the MRR_t of the MRSTP aligned well with the MRR_e. The average error was only 4.1%. The accuracy and validity of the MRR model was verified.

7.2 Surface morphology after polishing

Figure 17 shows the 3D morphology of the workpiece (Aluminum alloy 6061) surface before and after MRSTP. It was observed that there were obvious scratches on the workpiece surface before polishing. After polishing for 1 h, the scratches on the workpiece surface were completely removed. The surface roughness (S_a) of the workpiece surface decreased from 337.6 nm to 79.0 nm, which improved by 77%. The results demonstrated that MRSTP was effective to obtain excellent polishing performance for the Aluminum alloy 6061 workpiece.

8 Conclusions

In this work, CFD method was employed to investigate the effect of polishing parameters on normal and shear stress in the MRSTP process. The MRR model of MRSTP process was established based on the CFD pressure output and the Preston equation. The accuracy and validity of the MRR model was verified. The effect of polishing parameters on the MRR was investigated. The main conclusions were drawn as follows:

(1). The CFD simulation results showed that the pressure on the workpiece surface decreased with the increase of the rotational speed of the workpiece while it increased with the increase of the flow velocity of the MRSTP media. A continuous increment in shear stress on the workpiece surface was observed with the increase of the rotation speed of the workpiece and the flow velocity of the MRSTP media.

(2). The MRR model indicated that the MRR had relation with the MRSTP geometry configuration and process variables. The key process variables were the rotational speeds of the spindle and rotary table and working gaps.

(3). The accuracy and validity of the simulated results and MRR model were verified. The average error between theoretical and experimental results was 4.1%.

(4). The workpiece surface roughness (S_a) decreased to 79.0 nm from initial 337.6 nm with 77% improvement after MRSTP. The scratches on the workpiece surface were completely removed after polishing for 1 h.

Author contributions Zhen Ma: Writing-original draft, Methodology, Experiment, Investigation, Yebing Tian: Conceptualization, Supervision, Investigation, Writing-review and editing, Project administration, Funding acquisition, Cheng Qian: Experiment, Investigation, Shadab Ahmad: Writing-review, Zenghua Fan: Writing-review, Zhiguang Sun: Experiment.

Funding This work was financially supported by the National Natural Science Foundation of China (Grant No. 51875329), Taishan Scholar Special Foundation of Shandong Province (Grant No. tsqn201812064), Shandong Provincial Natural Science Foundation, China (Grant No. ZR2023ME112, ZR2017MEE050), Shandong Provincial Key Research and Development Project, China (Grant No. 2018GGX103008), Scientific Innovation Project for Young Scientists in Shandong Provincial Universities (Grant No. 2019KJB030), and Key Research and Development Project of Zibo City (Grant No. 2019ZBXC070).

Data availability All authors confirm that all data and materials reported in this paper are available.

Code availability Not applicable.

Declarations

Ethics approval This study complies with the ethical standards set out by Springer. All authors read and approve the final manuscript.

Consent to participate Not applicable.

Consent for publication The manuscript is approved by all authors for publication.

Conflict of interest All authors declare that no conflict of interest exists.

References

1. Barman A, Das M (2018) Nano-finishing of bio-titanium alloy to generate different surface morphologies by changing magnetorheological polishing fluid compositions. Precis Eng 51:145–152. https://doi.org/10.1016/j.precisioneng.2017.08.003
2. Zhai Q, Zhai WJ, Gao B (2021) Modeling of forces and material removal rate in ultrasound assisted magnetorheological polishing (UAMP) of sapphire. Colloids Surf, A 628:127272. https://doi.org/10.1016/j.colsurfa.2021.127272
3. Luo H, Guo MJ, Yin SH, Chen FJ, Huang S, Lu AG, Guo YF (2018) An atomic-scale and high efficiency finishing method of zirconia ceramics by using magnetorheological finishing. Appl Surf Sci 444:569–577. https://doi.org/10.1016/j.apsusc.2018.03.091
4. Yamaguchi H, Srivastava A, Tan M, Hashimoto F (2014) Magnetic Abrasive Finishing of cutting tools for high-speed machining of titanium alloys. CIRP J Manuf Sci Tec 7(4):299–304. https://doi.org/10.1016/j.cirpj.2014.08.002
5. Li K, Ma RJ, Zhang M, Chen W, Li XB, Zhang DZ, Tang Q, Murr LE, Li JF, Cao HJ (2022) Hybrid post-processing effects of magnetic abrasive finishing and heat treatment on surface integrity and mechanical properties of additively manufactured Inconel 718 superalloys. J Mater Sci Technol 128:10–21. https://doi.org/10.1016/j.jmst.2022.03.026
6. Umehara N, Kirtane T, Gerlick R, Jain VK, Komanduri R (2006) A new apparatus for finishing large size/large batch silicon nitride (Si3N4) balls for hybrid bearing applications by magnetic float polishing (MFP). Int J Mach Tool Manu 46(2):151–169. https://doi.org/10.1016/j.ijmachtools.2005.04.015
7. Kumar M, Bharti RK, Dass M (2022) Study of surface finishing mechanism in a rotational-magnetorheological miniature gear profile polishing using novel flow restrictor. Wear 488–489:204120. https://doi.org/10.1016/j.wear.2021.204120

8. Wang CJ, Cheung CF, Ho LT, Yung KL, Kong LB (2020) A novel magnetic field-assisted mass polishing of freeform surfaces. J Mater Process Tech 279:116552. https://doi.org/10.1016/j.jmatprotec.2019.116552
9. Liu JB, Li XY, Zhang YF, Tian D, Ye MH, Wang C (2020) Predicting the Material Removal Rate (MRR) in surface Magnetorheological Finishing (MRF) based on the synergistic effect of pressure and shear stress. Appl Surf Sci 504:144492. https://doi.org/10.1016/j.apsusc.2019.144492
10. Sharma VK (2021) Modeling and analysis of a novel rotational magnetorheological abrasive flow finishing process. Int J Lightweight Mater Manuf 4(3):290–301. https://doi.org/10.1016/j.ijlmm.2021.02.001
11. Grover V, Singh AK (2018) Modelling of surface roughness in a new magnetorheological honing process for internal finishing of cylindrical workpieces. Int J Mech Sci 144:679–695. https://doi.org/10.1016/j.ijmecsci.2018.05.058
12. Guo YF, Yin SH, Ohmori H, Li M, Chen FJ (2022) A novel high efficiency magnetorheological polishing process excited by Halbach array magnetic field. Precis Eng 74:175–185. https://doi.org/10.1016/j.precisioneng.2021.11.011
13. Chen FJ, Yin SH, Zhu KJ, Ohmori H, Zhu YJ, Fan YF (2009) Two-dimension modelling for material removal of magneto-rheological finishing. China Mech Eng 20(14):1647–1650
14. Ming Y, Huang X, Zhou D, Li X (2022) A novel Non-Newtonian fluid polishing technique for zirconia ceramics based on the weak magnetorheological strengthening thickening effect. Ceram Int 48(5):7192–7203. https://doi.org/10.1016/j.ceramint.2021.11.280
15. Li WH, Li XH, Yang SQ, Li WD (2018) A newly developed media for magnetic abrasive finishing process: Material removal behavior and finishing performance. J Mater Process Tech 260:20–29. https://doi.org/10.1016/j.jmatprotec.2018.05.007
16. Qian C, Tian YB, Fan ZH, Sun ZG, Ma Z (2022) Investigation on rheological characteristics of magnetorheological shear thickening fluids mixed with micro CBN abrasive particles. Smart Mater Struct 31:095004. https://doi.org/10.1088/1361-665X/ac7bbd
17. Fan ZH, Tian YB, Zhou Q, Shi C (2020) Enhanced magnetic abrasive finishing of Ti-6Al-4V using shear thickening fluids additives. Precis Eng 64:300–306. https://doi.org/10.1016/j.precisioneng.2020.05.001
18. Zhou Q, Tian YB, Fan ZH, Qian C, Sun ZG (2021) Study on Preparation and Processing Characteristics of Magnetic Shear Thickening Finishing Media. Surf Technol 50(7):367–375. https://doi.org/10.16490/j.cnki.issn.1001-3660.2021.07.039
19. Sun ZG, Fan ZH, Tian YB, Prakash C, Guo J, Li L (2022) Post-processing of additively manufactured microstructures using alternating-magnetic field-assisted finishing. J Mater Res Technol 19:1922–1933. https://doi.org/10.1016/j.jmrt.2022.05.176
20. Chen FJ, Hao SM, Miao XL, Yin SH, Huang S (2018) Numerical and experimental study on low-pressure abrasive flow polishing of rectangular microgroove. Powder Technol 327:215–222. https://doi.org/10.1016/j.powtec.2017.12.062
21. Li M, Lyu BH, Yuan JL, Dong CC, Dai WT (2015) Shear-thickening polishing method. Int J Mach Tool Manu 94:88–99. https://doi.org/10.1016/j.ijmachtools.2015.04.010
22. Li M, Liu MH, Riemer O, Karpuschewski B, Tang C (2021) Origin of material removal mechanism in shear thickening-chemical polishing. Int J Mach Tool Manu 170:103800. https://doi.org/10.1016/j.ijmachtools.2021.103800
23. Li M, Karpuschewski B, Riemer O (2022) Controllable shearing chain-thickening polishing process for machining of barium borate. Int J Mach Tool Manu 37:291–301. https://doi.org/10.1016/j.cirpj.2022.02.005
24. Zhu WL, Beaucamp A (2022) Generic three-dimensional model of freeform surface polishing with non-Newtonian fluids. Int J Mach Tool Manu 172:103837. https://doi.org/10.1016/j.ijmachtools.2021.103837
25. Kumar M, Ahmad S, Das M (2021) Magnetorheological finishing of miniature gear teeth profiles using uniform flow restrictor. Mater Manuf Processes 37(4):467–482. https://doi.org/10.1080/10426914.2021.1954193
26. Prakash C, Singh S, Pramanik A, Basak A, Królczyk G, Bogdan-Chudy M, Wu YL, Zheng HY (2021) Experimental investigation into nano-finishing of β-TNTZ alloy using magnetorheological fluid magnetic abrasive finishing process for orthopedic applications. J Mater Res Technol 11:600–617. https://doi.org/10.1016/j.jmrt.2021.01.046
27. Preston FW (1927) The Theory and Design of Plate Glass Polishing Machines. J Soc Glass Technol 11:214–256
28. Zhu WL, Beaucamp A (2020) Non-Newtonian fluid based contactless sub-aperture polishing. CIRP Ann-Manuf Techn 69(1):293–296. https://doi.org/10.1016/j.cirp.2020.04.093
29. Deng JY, Zhang QX, Lu JB, Yan QS, Pan JS, Chen R (2021) Prediction of the surface roughness and material removal rate in chemical mechanical polishing of single-crystal SiC via a back-propagation neural network. Precis Eng 72:102–110. https://doi.org/10.1016/j.precisioneng.2021.04.012
30. Wang QD, Liu MC, Zhang HX (2000) Removing function of polishing pad in computer controlled optical polishing. Opt Tech 26(1):32–34. https://doi.org/10.13741/j.cnki.11-879/o4.2000.01.010
31. Tian YB, Shi C, Fan ZH, Zhou Q (2020) Experimental investigations on magnetic abrasive finishing of Ti-6Al-4V using a multiple pole-tip finishing tool. Int J Adv Manuf Tech 106:3071–3080. https://doi.org/10.1007/s00170-019-04871-z
32. Zhou DD, Huang XM, Ming Y, Li XY, Li HY, Li W (2021) Material removal characteristics of magnetic-field enhanced shear thickening polishing technology. J Mater Res Technol 15:2697–2710. https://doi.org/10.1016/j.jmrt.2021.09.092

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Check for updates

Modified Decoupled Sense Amplifier with Improved Sensing Speed for Low-Voltage Differential SRAM

AYUSH, Delhi Technological University, India
POORNIMA MITTAL, Delhi Technological University, India
RAJESH ROHILLA, Delhi Technological University, India

A modified decoupled sense amplifier (MDSA) and modified decoupled sense amplifier with NMOS foot-switch is proposed for improved sensing in differential SRAM for low voltage operation at 22 nm technology node. The MDSA and MDSANF both offer notable improvements to read delay over conventional voltage and current sense amplifiers. At an operating voltage of 0.8 V, the MDSA exhibited a reduced delay of 28.6%, 41.79%, 37.74%, 30.94% compared to modified clamped sense amplifier (MCSA), double tail sense amplifier(DTSA), modified hybrid sense amplifier (MHSA) and conventional latch-type sense amplifier (LSA) respectively. Similarly, the MDSANF demonstrated a delay reduction of 26.13%, 39.78%, 35.58%, 28.55% over MCSA, DTSA, MHSA and LSA respectively. To validate the performance, the MDSA and MDSANF are evaluated using the variation in delay and power consumption across various supply voltages, process corners, input differential bit line voltage (ΔV_{BL}), bit line capacitance C_{BL}) and the sizing of decoupling transistors. Monte Carlo simulations were conducted to analyse the impact of voltage threshold variations on transistor mismatch which leads to an increased occurrence of read failures and a decline in SRAM yield. The performance analysis of various voltage and current sense amplifiers is presented along with MDSA and MDSANF. Area consideration for selection of sensing scheme is important and as such layout of MDSA and MDSANF was performed conforming to the design rules and estimated area for MDSA is 0.297 μm^2 whereas MDSANF occupies 0.5192 μm^2.

CCS Concepts: • **Hardware → Very large scale integration design**; **Process variations**.

Additional Key Words and Phrases: low voltage SRAM, sensing delay, sense amplifier, SRAM sensing, transient analysis

1 INTRODUCTION

The backbone of numerous system-on-chip (SoC) devices used in crucial industries like the Internet of Things (IoT), biomedical implants, and wearable smart devices is a fast sensing, low-energy, and reliable Static Random Access Memory (SRAM). The SRAM is typically the dominant contributor to power consumption as well area occupied of the whole chip [Zhu and Kursun 2014]. This puts the stability, delay, yield and energy consumption of SRAM at the forefront of the design strategy for high performance systems. As a result of the stringent design parameters for SRAM, the selection of bit cells, such as 6T, 7T [Rawat and Mittal 2021, 2022a,b,c], 8T [Lorenzo and Paily 2022], 9T [Abbasian and Gholipour 2022] etc. together with peripheral circuitry including address decoder, column multiplexer [Mittal et al. 2022], sense amplifier (SA) and write driver plays a crucial role in the design process. The ability to operate the system at lower voltages is one of the approaches employed in order to decrease the energy consumption thereby increasing the battery's life cycle. However, The operation of the SRAM at the minimum supply voltages with the minimum sized transistors for sub-45 nm technologies introduces various process, voltage and temperature (PVT) variations which when compared with operation at

Just Accepted

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
1084-4309/2023/8-ART $15.00
https://doi.org/10.1145/3611672

nominal supply voltages and larger device sizing, could result in a two-fold speed degradation [Reyserhove and Dehaene 2017].

The Sense Amplifier (SA) plays a significant role in the read operation of SRAM. The SA performs the amplification of a small differential signal present at the bit lines to larger rail-to-rail swing although single-ended SA architectures have also been adopted in SRAM [Rawat and Mittal 2022d; Shen et al. 2022]. Since, SRAM cells are highly sensitive to process variations, the SA is also affected by the random variations caused by Random Dopant Fluctuations (RDF). The mismatch in threshold voltages has a vital impact in deteriorating the SRAM read yield and the access delay.

The SA topologies determine the performance and operation limits of the SRAM. The widely adopted voltage-mode SAs are based on cross-coupled inverter latches, which are simple and robust [Dutt et al. 2022; Kobayashi et al. 1993; Schinkel et al. 2007; Wicht et al. 2004]. SAs have a small delay due to the positive feedback of cross-coupled inverters, while voltage-mode SAs require a significant voltage difference between a pair of bit lines due to their offset. Large delays result from the small bit cell driving large capacitive bit lines. Bit lines can be reduced in capacitance by using divided bit lines and two-level sensing [Schinkel et al. 2007; Wicht et al. 2004]. Since the local bit line connects fewer bit cells, there is a lower bit line discharge delay, but the bit line full voltage swing consumes more energy.

Current-mode SAs may have more advantages because of no extra intermediate voltage [Blalock and Jaeger 1991; Chee et al. 1992; Do et al. 2007; Khellah and Elmasry 2001; Wang and Lee 1998]. The current conveyor widely utilized in current-mode SAs uses four PMOS transistors to create a virtual short circuit through the bit lines as it reduces the sensing delay and is not immune to bit line capacitance variations [Seevinck et al. 1991]. Also, a clamped bit line current-mode SA with low impedance was presented [Blalock and Jaeger 1991], which maintains a constant potential on the bit lines. Following this, most research on current-mode SAs has focused on circuit improvement based on the current conveyor and the clamped bit line technique to improve speed and energy efficiency.

The process variation has a large effect on SA offset with CMOS technology node scaling. SA offset caused by inherent mismatch at lower technology nodes has a negative impact on SRAM yield. Many solutions have been reported to reduce offset [Bhavnagarwala et al. 2001; Fragasse et al. 2019; Patel et al. 2021; Reniwal et al. 2019; Shah et al. 2013; Zhao et al. 2022] with one of the most basic approaches being the use of enlarged transistors [Bhavnagarwala et al. 2001]. Multiple redundant SAs of same [Verma and Chandrakasan 2007] and different sizes [Sharma et al. 2011] were also used together with advanced calibration schemes to counter offset. These calibration-based SA topologies allow for high speed operation with lower power consumption while having a larger area overhead and increased complexity.

In this paper, we propose a modified decoupled sense amplifier (MDSA) and a modified decoupled sense amplifier with NMOS foot-switch for performing differential-ended SRAM cell read operations. The Section 2 consists of the conventional SAs being classified as either voltage or current sense amplifiers. The operation and structure of MDSA and MDSANF are discussed in Section 3 together with the impact of decoupling PMOS sizing on read performance as well as reliability of the SA under parameter variations. The Section 4 consists of performance analyis based on simulation results of the conventional voltage and current SAs with MDSA and MDSANF across the set of supply voltages. The work in concluded in Section 5.

2 CONVENTIONAL SENSE AMPLIFIER ARCHITECTURES

2.1 Voltage Sense Amplifiers

Voltage-mode sense amplifiers are composed primarily of a cross-coupled inverter latch which exhibits a strong positive feedback. The latch-type sense amplifier (LSA) is shown in Figure 1a. is one of the earliest forms of CMOS-based sense amplifier used in memory design. The cross-coupled inverters are activated using the sense

Just Accepted

enable (SA_EN) signal, which turns on transistor M7. The circuit is elementary and robust in its operation; however, owing to its large bit line capacitance, this SA exhibits increased delay and power consumption. To limit the power consumption and decrease delay, transistors M3 and M4 are added to the LSA to obtain the current-controlled latch sense amplifier (CLSA)[Kobayashi et al. 1993]. Its circuit schematic is depicted in Figure 1b. The activation of transistor M7 is performed using the enable signal (SA_EN), which lends itself to the current flow through M3 and M4. The high input impedance created at the bit lines allows faster sensing. Pass transistors can be used to mitigate the large capacitive bit lines when connected to bigger cell arrays.

Two bidirectional pass transistors, M5 and M6, are used in the decoupled latch-type sense amplifier (DLSA) [Wicht et al. 2004] in Figure 1c. These transistors decouple the bit lines so that cross-coupled inverters are no longer used to drive the massive capacitive bit lines, which speeds sensing and reduces power consumption. In order to ensure proper operation for DLSA, the per-charge voltage (VPC) is used to per-charge the bit lines to V_{DD}.

The double-tail sense amplifier (DTSA) [Schinkel et al. 2007] illustrated in Figure 1d is a sensing technique which uses a two tail approach, one focused on the input bit lines and the other on the cross-coupled inverter latch. During the pre-charge phase, the nodes Di+ and Di- of PMOS transistors M8 and M9 are charged to V_{DD}. This causes transistors M5 and M6 to pull the output nodes (SAO+ and SAO-) to GND. When the clock signal (CLK) goes high, both M7 and M12 transistors are activated, leading to a gradual decline in the voltages at the Di nodes. The rate of voltage decrease is determined by the current flowing through M12 and the capacitance at the Di nodes. This generates an input-dependent voltage difference, denoted as ΔVDi. The cross-coupled inverter pair receives this voltage difference, regenerating it when the SAO nodes can no longer be clamped to ground by M5 and M6. Consequently, the DTSA offers effective isolation between the input and output signals, thanks to M5 and M6, and it can operate well with lower supply voltages.

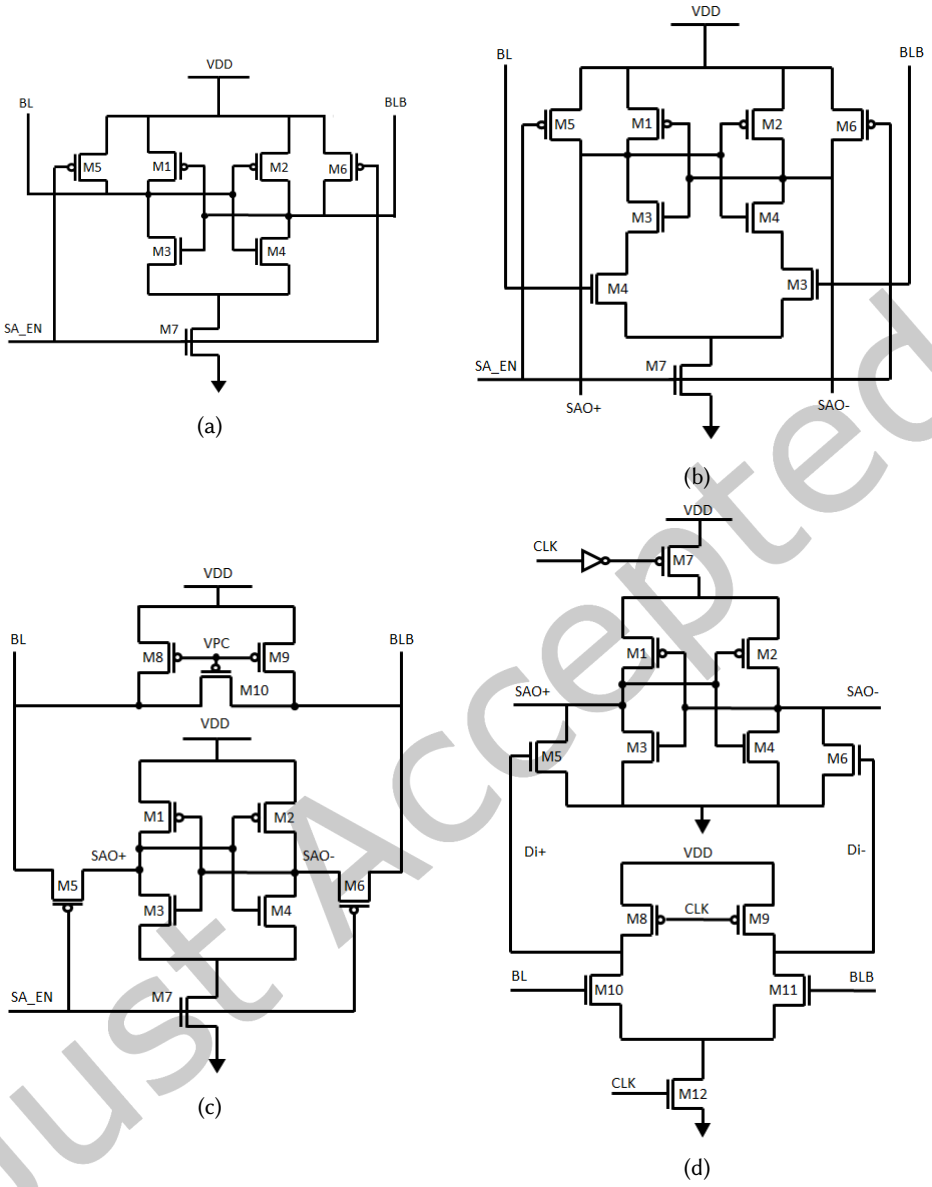
2.2 Current Sense Amplifiers

The read operation of the current sense amplifiers is independent of the differential voltage present on the bit lines, making them a popular choice for faster and larger SRAM systems. The clamped bit line sense amplifier (CBLSA) [Blalock and Jaeger 1991] shown in Figure 2a which uses the transistors M8 and M9 as a clamp for the bit lines. This SA works in a two-stage format with the first stage being per-charge made up of the transistor pair M6 and M7. Subsequently, to start the read operation for the memory, the large-width M1 transistor was turned on, which enabled the SA. The M6 and M7 transistors works during the per-charge phase to equalize the bit line voltages. The current through the SA is a design outcome of the sizing of M2-M5. This current value is chosen to produce a voltage difference at the bit lines of around 100mV. The per-charge stage is completed when the per-charge signal (VPC) is low, thereby rendering the cross-coupled inverter pair to act as a high-gain amplifying stage. The voltage difference owing to the differential current is amplified, which leads to a large potential difference at the output nodes SAO+ and SAO-.

The hybrid current-mode sense amplifier (HCSA) [Chee et al. 1992] is shown in Figure 2b. It consists of four PMOS transistors of equal size, M10–M13, which together form the current conveyer [Seevinck et al. 1991]. This just serves as a current transport mechanism without introducing any gain. It is possible to think of the HCSA as a CBLSA with a current conveyer. The HCSA's read procedure is as follows: SA_EN = 1 activates M7, which causes a differential current signal to arise at bit lines BL and BLB. The differential voltage generated at the drains of The NMOS devices M3 and M4 is due to the differential current signal at their source terminals. This modest differential voltage is subsequently amplified by the cross-coupled latch to create a larger voltage swing.

The design shown in Figure 2c is a modified clamped sense amplifier (MCSA) [Khellah and Elmasry 2001]. The structure is comprised of CBLSA and equalizing transistor M7. The two-stage read operation works as follows. The bit lines are pre-charged by using transistor M7 controlled by the SA_EN going high. Initially, bit lines are

Just Accepted



Just Accepted

Fig. 1. Circuit schematics for Voltage SAs (a) Conventional latch-type SA, (b) Current-controlled latch-type SA, (c) Decoupled latch-type SA, and (d) Double tail SA

equalized by switching the SA_EN high to turn on M7. During the sensing stage, SA_EN was turned off. The operation of the MCSA is governed by the same steps as the CBLSA, with the quiescent differential voltage generated at the bit lines being tracked and amplified. Figure 2d presents the modified hybrid sense amplifier (MHSA) [Do et al. 2007] which is comprised of the conventional current conveyer together with YSEL operated

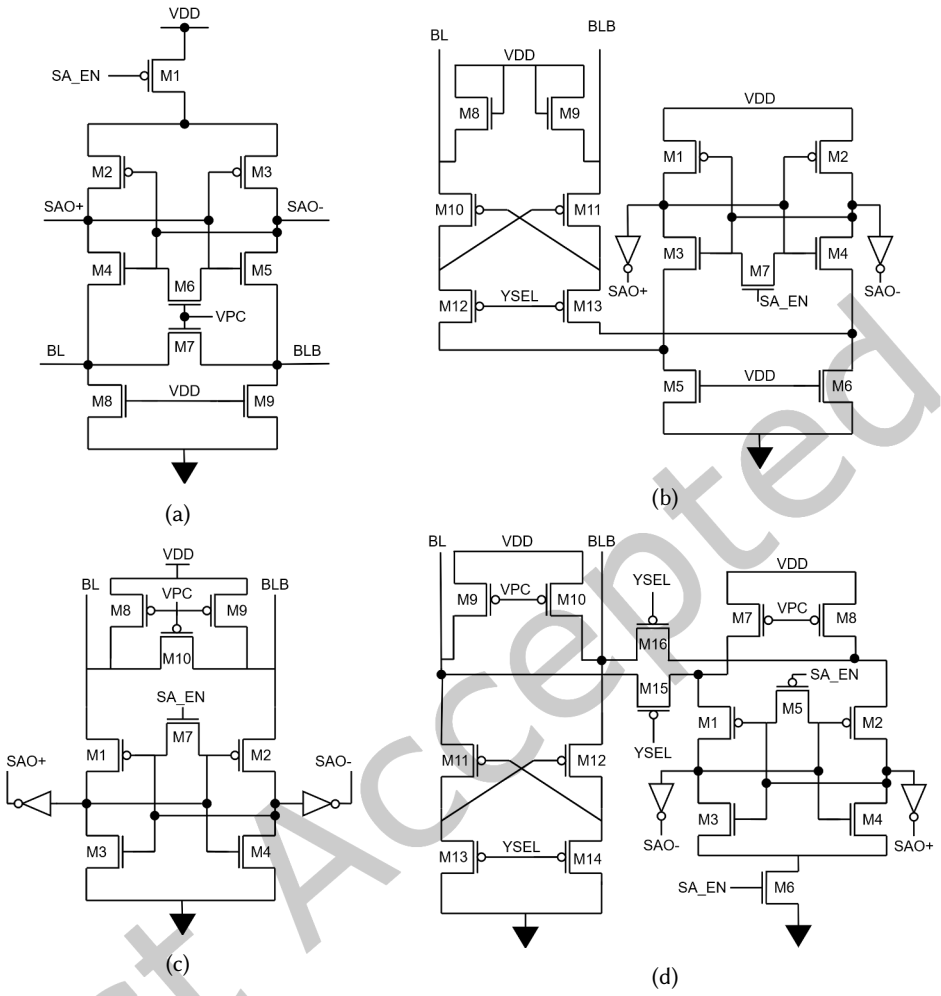


Fig. 2. Circuit schematics for Current SAs (a) Clamped bit line SA, (b) Hybrid current SA, (c) Modified clamped SA, and (d) Modified hybrid current SA

transistors to couple the bit lines and the data lines. The per-charge stage consists of transistors M9 and M10 which pulls the bit lines BL and BLB high. The rest of the sense amplifier working is similar to the CBLSA with the small potential difference at the bit lines being amplified to produce the read output.

3 MODIFIED DECOUPLED SENSE AMPLIFIER

3.1 Design Architecture and Operation of proposed SA

Figure 3a and 3b illustrates the detailed circuit schematic of the proposed modified decoupled sense amplifier (MDSA) and MDSA with NMOS footswitch (MDSANF) respectively. Both the SA architectures are designed using the conventional inverter latch decoupled by the PMOS switches M6 and M7. The operation of the SAs can be divided into three segments. The three segments are the pre-charge, amplification and output propagation. The

Just Accepted

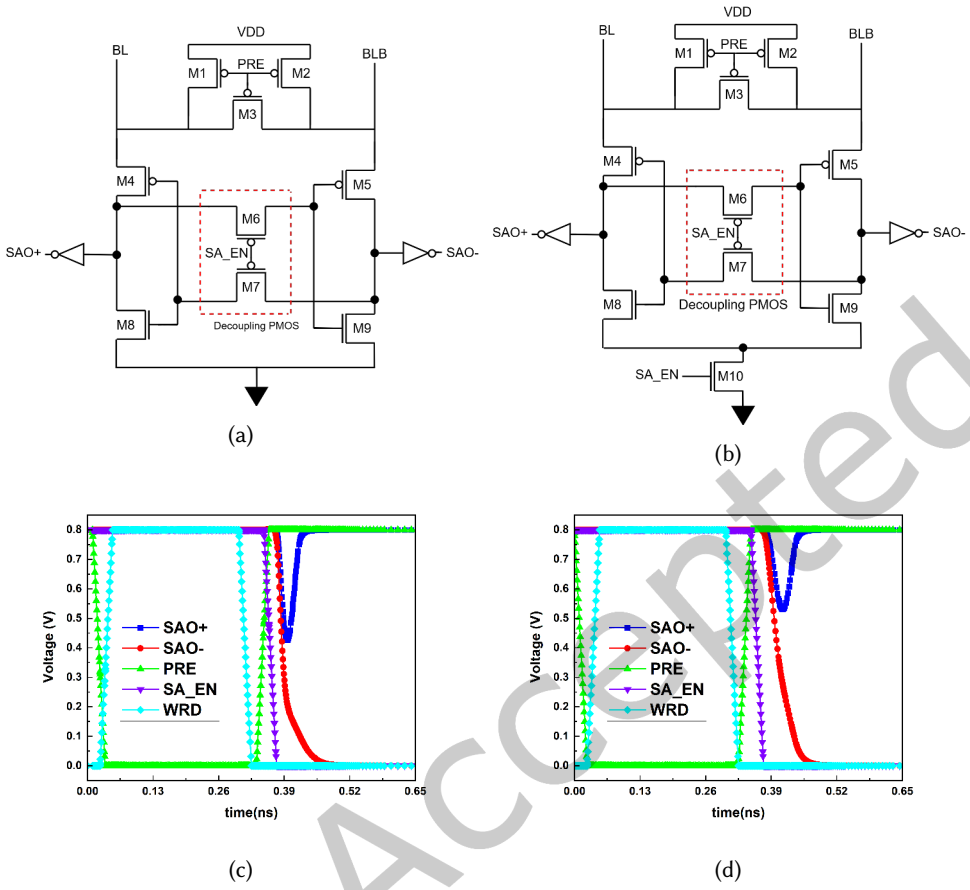


Fig. 3. Circuit schematics for (a) MDSA and (b) MDSANF, along with the transient response showing timing control signals at V_{DD} = 0.8 V and C_{BL} = 50 fF for (c) MDSA and (d) MDSANF

pre-charge segment requires the bit lines to be charged to supply voltage V_{DD} using M1-M3 PMOS transistors. The timing control signal (TCS) used here is PRE. The amplification stage is initiated by turning on decoupling switches (M6 and M7) by turning the TCS SA_EN off. The cross-coupled latch is comprised of two inverters formed by the M5, M8 and M5, M9 transistors respectively. The latch then acts as a high-gain positive feedback amplifier wherein a differential current generated across the drains of the M8 and M9 gives rise to large differential voltage across SAO+ and SAO-. The SAO+ is pulled high while the SAO- is pulled to GND for a read "1" operation and vice versa. The outputs nodes of the SA are passed through inverters to attain better isolation and the maximum output signal swing. This contributes to lower delay during the read operation and improves the performance of the sense amplifier. The additional NMOS footswitch M10 used is similar to the ones used in conventional voltage-mode SAs. It serves as a sleep transistor to reduce leakage in the power gating aspect of the SA. The figure 3c and 3d depict the transient response for MDSA and MDSANF respectively at V_{DD} = 0.8 V with C_{BL}(bit line capacitance) being 50fF. The transient response is used to derive the delay performance of the SA. The delay of the sense amplifier assumed to be the time between the 50% point of the SRAM word line (WL) and the 50% point of the differential voltage at SAO nodes [Chen et al. 2021].

Just Accepted

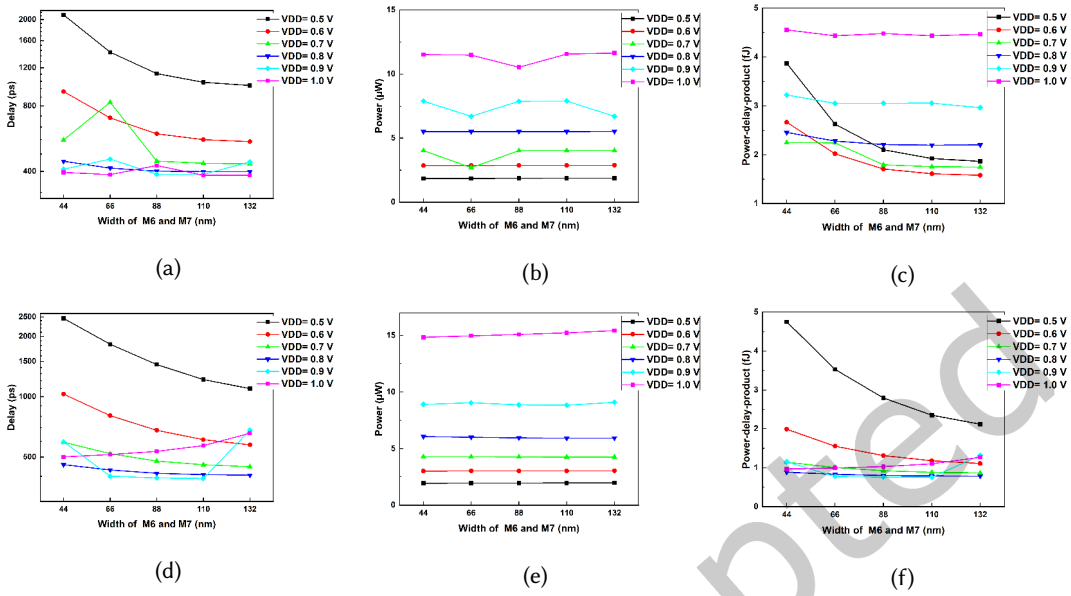


Fig. 4. Investigating the impact of PMOS devices M6 and M7 width on MDSA and MDSANF performance: (a) Delay, (b) Power consumption, and (c) Power-delay-product for MDSA and (d) Delay, (e) Power consumption, and (f) Power-delay-product of MDSANF examined for different supply voltages

3.2 Impact of Sizing of Decoupling PMOS on SA Read Performance

To decouple the inverters in the latch before the SA is enabled, the MDCA/MDSANF architecture uses the PMOS transistors M6 and M7. Figure 4 shows the impact of the variation in widths of PMOS transistors M6 and M7 on the read performance of the proposed SA. Figure 4a and 4b illustrate the delay variation of MDSA and MDSANF respectively for the given set of supply voltages. The delay is lowest for 1 V supply and is consistently larger than 1 ns for all widths for near threshold operation at 0.5 V. The widths of the PMOS transistors have a minimal effect on the power consumption at all supply voltages except 1 V as is evident in figure 4b and 4d. The largest transistor width which performs a correct read operation for the value stored in the 6T bit cell is 88 nm. This width is considered for all best-case calculation from here onwards. This is done so as to minimize delay while performing a reliable read operation across the set of supply voltages for both the MDSA and MDSANF. The transistors of the inverter latch are designed with PMOS/NMOS ratio of 2 with minimum sized NMOS transistors. The pre-charge circuit made up of transistors M1-M3 are sized at larger W/L = 1000 nm/22 nm. The process variation tolerance is also improved by utilizing the largest width M6 and M6 at 88 nm which allows a successful SRAM read.

3.3 Parameter Variation Tolerance of Proposed SA

CMOS parameter variations can be classified into two main categories: front-end variations and back-end variations. Among these, front-end random variations are primarily influenced by the fluctuation in threshold voltage (VTH). One of the key contributors to front-end random variations is random dopant fluctuations (RDF). RDF refers to the variations in the number of dopant atoms present in the transistor channel. As technology nodes decrease, the number of dopant atoms decreases significantly, reaching around ten atoms at 16 nm compared to

Just Accepted

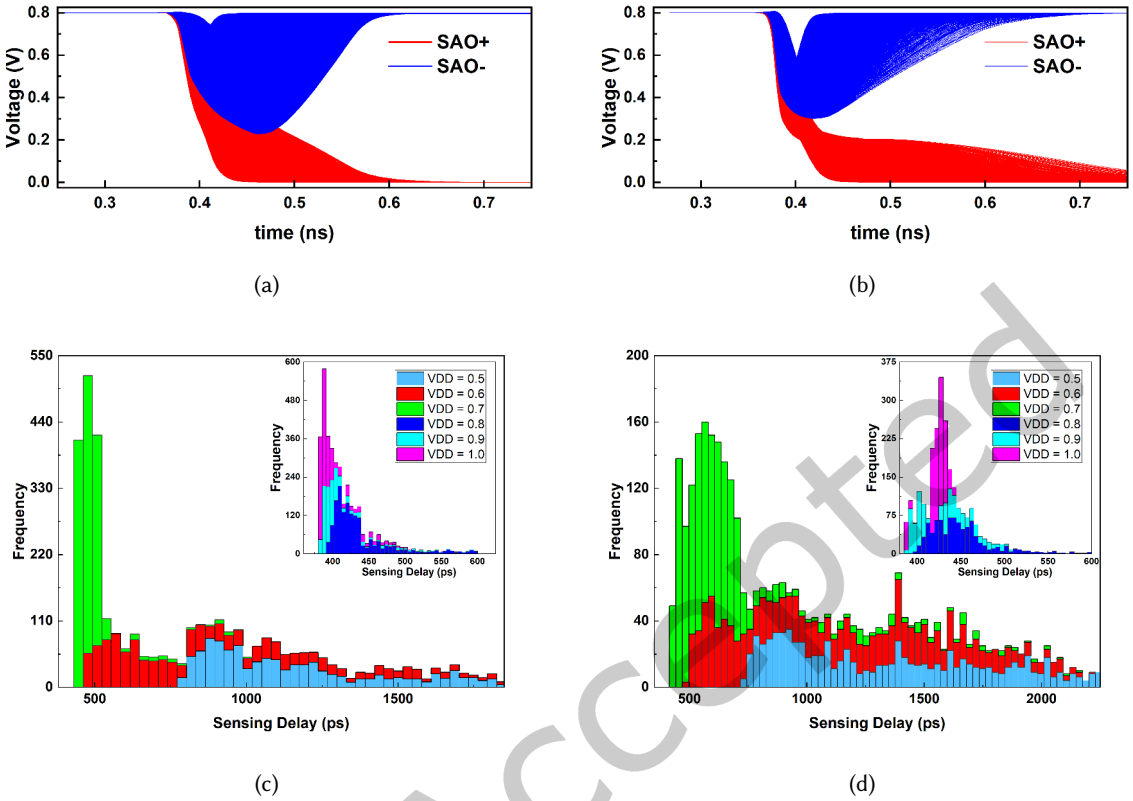


Fig. 5. Transient response (10k Monte Carlo simulation) of (a) MDSA, and (b) MDSANF and calculated delay distribution at C_{BL} = 50fF for near threshold operation of (c) MDSA and (d) MDSANF (Inset: delay distribution at above threshold operation)

a few tens at 45 nm [Hernández et al. 2011]. This reduction in dopant atoms contributes to the inconsistency in transistor parameters, thereby increasing the variability in CMOS devices at smaller technology nodes.The effect of process variation on the MDSA and MDSANF is studied using Monte Carlo simulations for 10k data points. A read failure might happen as a result of V_{TH} fluctuations, which is defined as the read output for a read "1" operation going to zero rather than V_{DD}.The transient responses at V_{DD} = 0.8 V and C_{BL} = 50 fF for MDSA and MDSANF are depicted in Figure 5a and 5b. The transient response depicts that there are zero read failures that take place during the SRAM read. It is assumed that the threshold voltage fluctuation follows a Gaussian distribution with 3σ variation around the mean μ. The lowest read failures are used to determine a SA's yield. A successful read operation in SRAM critically dependent on the symmetry of the majority of differential-ended SAs. The distribution of read delay for MDSA is shown in Figure 5c for near threshold operation and Figure 5c (inset) for other supply voltages. Similarly, the delay distribution for MDSANF depicted in Figure 5d and 5d (inset). The delay distributions are calculated from the MC simulations. The near threshold voltage operation shows deviation larger than a few hundred ps but goes down to tens of ps for higher supply voltages as is evident from Figure 5c and 5d for both MDSA and MDSANF.

To perform a performance analysis of a sense amplifier in an SRAM array, various design parameters can be considered, including the bit line capacitance. The bit line capacitance, ranging from 10 fF to 500 fF, is a crucial

Just Accepted

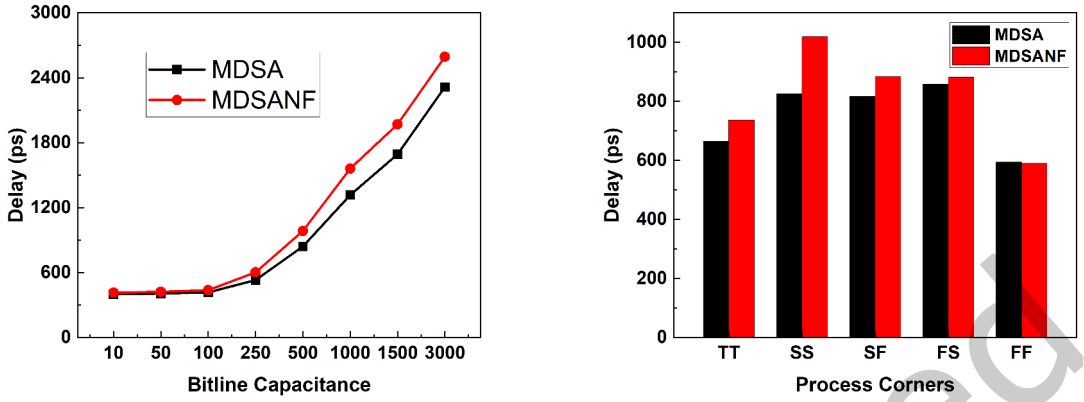


Fig. 6. Delay variation as a function of (a) Bitline Capacitance and (b) Process corners for MDSA and MDSANF at VDD = 0.8 V

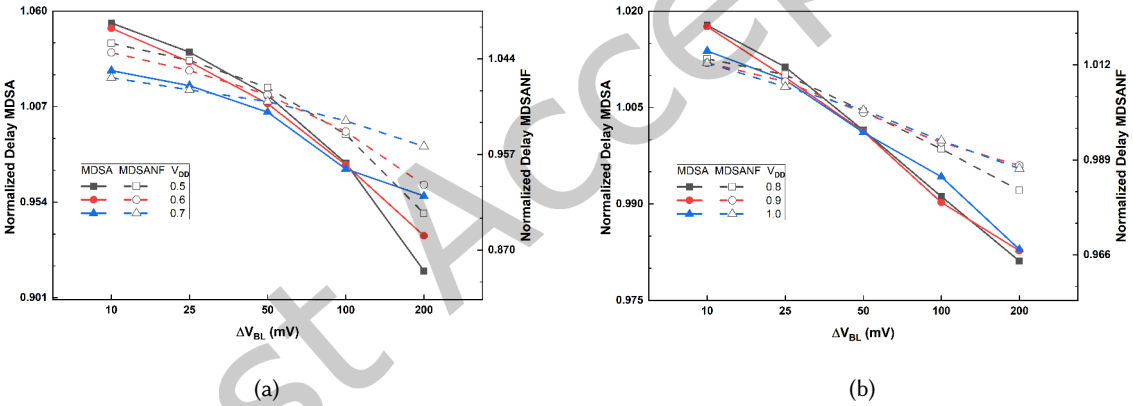


Fig. 7. Normalized Delay as a function of differential bit line voltage input for MDSA and MDSANF at (a) near threshold and (b) above threshold supply voltages

parameter that affects the speed and overall performance of the sense amplifier. By considering different values of bit line capacitance, representing varying array sizes, a comprehensive analysis can be conducted to assess the impact on the delay, power consumption, and noise margin of the sense amplifier in the SRAM array. Other design parameters that can be considered for performance analysis include the array size (number of memory cells), bit line capacitance, supply voltage, and process technology. Each of these parameters influences the behaviour of the sense amplifier and can be used to evaluate its performance under different operating conditions and array configurations. A singular SA is used to perform read operation for a column of bit cells which can range upwards of 1024 in large SRAM array structures. The delay variation as a function of C_{BL} is shown in Figure 6a. The increase in C_{BL} causes larger read delay due the SA being forced to discharge large capacitance to achieve

Just Accepted

a correct read. The delay shows a 4.77 times increase for MDSA for the bit line variation from 10 fF to 3000 fF and 6.25 times increase for MDSANF. To further justify the SA's robustness, the delay analysis for MDSA and MDSANF is extended across various global corners at V_{DD} = 0.8 V and C_{BL} = 50 fF and these results are shown in figure 6b. The normalized delay of the SAs as function of the input differential bit line voltage is shown in Figure 7 for the set of supply voltages. The normalized delay values show most deviation for near threshold operation of 0.5 V with deviation of 185.358 from fastest to slowest read delay and least deviation of only 10.932 ps at 1 V supply for MDSANF. Similarly, the largest delay deviation is 118.21 ps at 0.5 V while the smallest is 6.557 ps at 1 V for MDSA.

4 PERFORMANCE ANALYSIS OF MDSA/MDSANF WITH EXISTING SA ARCHITECTURES

The comparative analysis of all the conventional sense amplifier architectures is presented in this section with all the SAs being designed using Predictive Technology Model [Arizona State University and Group 2008] at 22 nm technology node. All the various SAs were designed using consistent sized transistors in order to conduct an objective analysis at various supply voltages. The conventional voltage and current-mode SAs have transistors with W/L = 44 nm/ 22 nm while all access transistors are sized with W/L = 1000 nm/ 22 nm. The read delay, power consumption and leakage have been calculated based on the simulation results. An important performance statistic for SA is the power-delay-product, or PDP for short. A stronger overall performance is shown by the lower PDP numbers. The identical word line (WL) and sense amplifier enable (SA_EN) timing signals are utilized while performing all simulations, together with the precise timing and control signals (TCS) as specified in the relevant literature. The simulation testbench schematic used to compare the sense amplifiers is shown in figure 8a. A single 6T SRAM cell is used together with a π-3 model for the bitline capacitance of large SRAM array structures. This is similar to test circuit in [Zhu et al. 2013].

4.1 Read Delay

The conventional SA architectures are simulated and the calculated read delay is shown in figure 8b. As is clearly evident, MDSA and MDSANF have the fastest operation for all supply voltages except the near threshold voltage operation at V_{DD} = 0.5 V. The proposed MDSA shows a 28.6% read delay improvement over MCSA while MDSANF shows a 26.13% improvement at V_{DD} = 0.8 V. The MCSA is the fastest among the voltage-mode and current-mode SAs followed by LSA and MHSA. MHSA, MCSA and DTSA exhibit the most consistent delay across voltage supply variation with DTSA having its best and worst delay 210.5 ps apart. The fastest operation at lowest applied supply is given by DTSA with MCSA being the second fastest as a result 10.49% increase in delay. . Amongst the current SAs compared, only CBLSA has delay consistently larger than 1 ns across all the supply voltages applied. The conventional SA designs compared above are all simulated for a bit line capacitance modelled by 50 fF with ΔV_{BL} assumed to be 50 mV with transistors having no mismatch operated at TT/27°C. The MDSA and MDSANF perform the best at voltage supply larger than 0.6 V and the delay improves as V_{DD} is increased. The MDSA and MDSANF have nearly same read delay at all voltages except 0.5 and 1 V where MDSA is faster by 22.05% and 20.32% respectively.

4.2 Power Consumption and Leakage Power

The power consumption of MDSA and MDSANF is compared to voltage-mode and current-mode SA architectures and the simulation results are shown in figure 9a. As the supply voltage increases, the power consumption of all SAs also increases, which aligns with the expected relationship between power, supply voltage, and current. The power consumption of the MDSA and MDSANF is moderate compared to other SA designs. Among the SA designs, the MHSA consistently exhibits the highest power consumption. This can be attributed to its larger number of transistors, particularly the current conveyor and data line coupling devices, which contribute significantly

Just Accepted

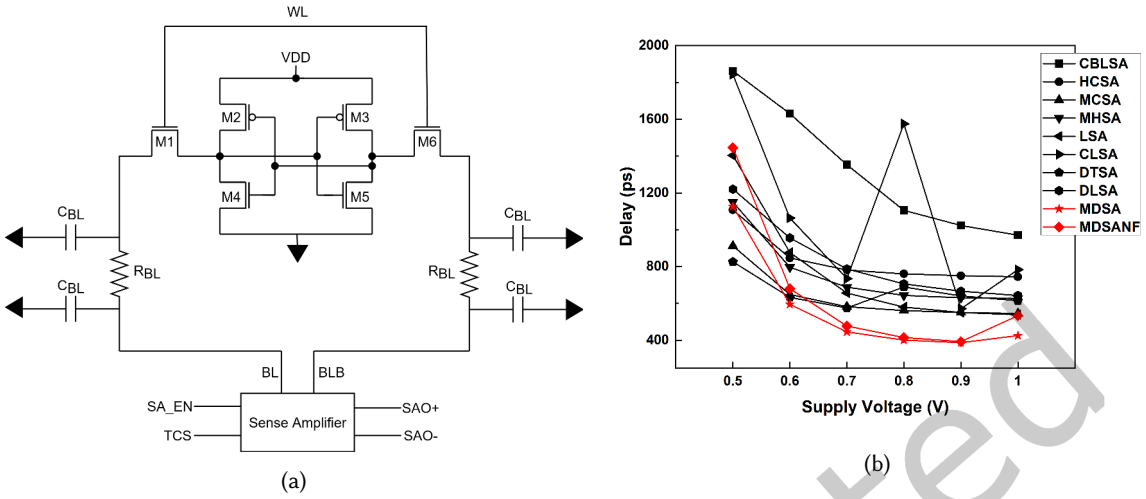


Fig. 8. (a) Simulation test-bench used for comparison of sense amplifiers (b) Delay comparison of all SAs as a function of the supply voltage (C_{BL} = 50 fF)

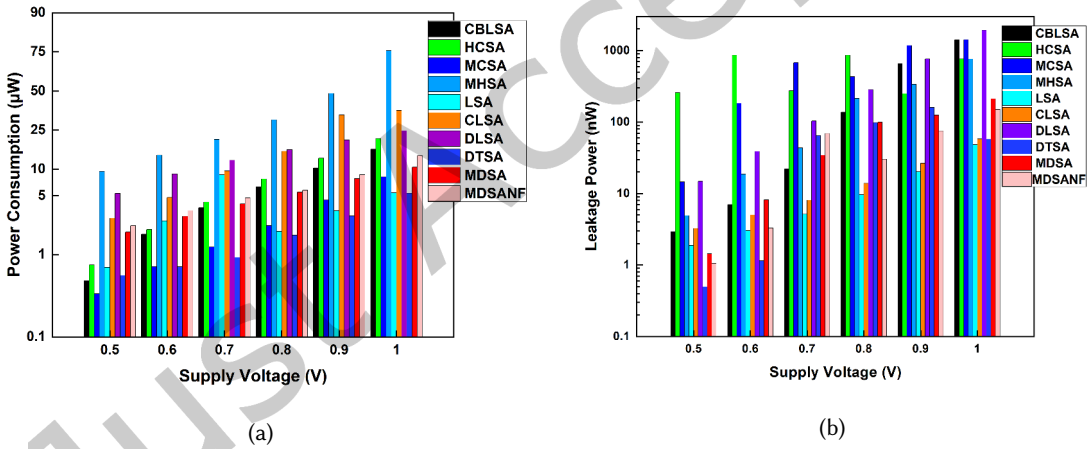


Fig. 9. (a) Power Consumption and (b) Leakage power comparison of all SAs as a function of the supply voltage (C_{BL} = 50 fF)

Just Accepted

to power consumption. The inclusion of additional control signals in the MHSA design further adds to the power consumption. In contrast, the CLSA and DLSA designs aim to minimize power consumption by reducing the charge and discharge of the bit line capacitance, resulting in lower power requirements. The presence of pre-charge bit lines also contributes significantly to power consumption in SA designs. The DTSA and MCSA have the lowest power consumption with averages of 2.86 μW and 2.03 μW respectively. When comparing to MCSA, MDSA shows a 2.30 μW increase while MDSANF shows an increment of 2.59 μW at V_{DD} = 0.8 V. However,

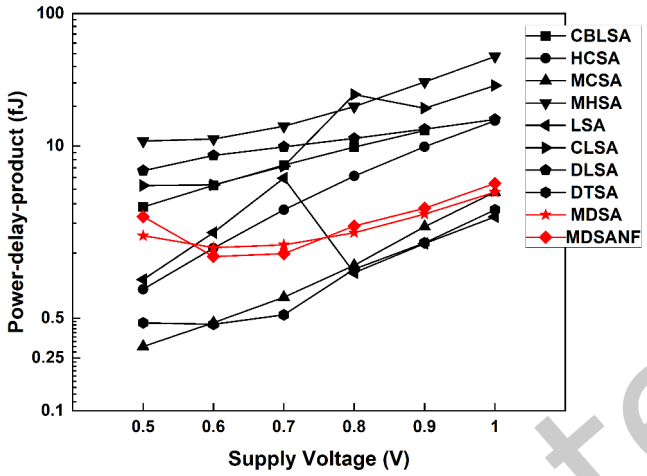


Fig. 10. Power-delay-product comparison of all SAs as a function of the supply voltage (C_{BL} = 50 fF)

the MDSA and MDSANF consume slightly higher power compared amongst the conventional SAs. Consequently, there is always a trade-off between sensing delay and power consumption when selecting the SA to perform SRAM read operation. Static power, which contributes significantly to the total power consumption of a circuit, is primarily caused by leakage current. It arises due to various factors such as sub-threshold current, gate leakage current, and reverse junction current in a MOS circuit. Mathematically, static power is determined by multiplying the supply voltage (VDD) with the leakage current.

Leakage current refers to the undesired flow of current in a transistor when it is in an idle or off state. The sub-threshold current occurs when there is a small voltage across the transistor, and it results from the imperfect blocking of current by the transistor's gate. The gate leakage current, on the other hand, happens due to the thin insulating layer of the transistor, which allows some current to pass through. Lastly, the reverse junction current occurs as a result of the inherent characteristics of the transistor's junctions. The leakage power for SAs is shown in figure 9b. The MDSANF utilizes the power gating effect of the NMOS foot-switch in order to decrease leakage power as compared to MDSA.

The Power-delay-product (PDP) of all SAs, assuming a bit line capacitance (C_{BL}) of 50 fF, is shown in figure 10. The MHSA consistently exhibits the highest PDP, with an average of 22.35 fJ. The proposed SA design shows a higher PDP than only for DTSA and MCSA, which are both characterized by their simple cross-coupled inverter-based architecture and the low transistor count. Therefore, the MDSA and MDSANF can be employed in faster sensing applications, albeit with a slightly higher power consumption.

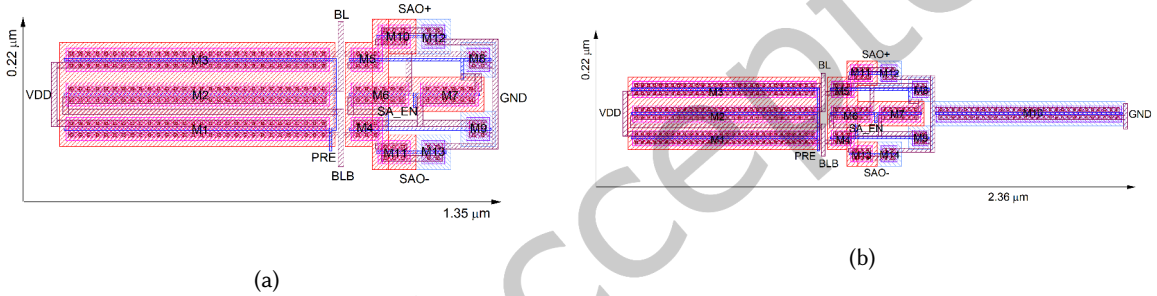
4.3 Area Considerations

The selection of a differential SA topology for performing SRAM read is a critical step as a single SA is used to read the data for a column of bit cells. Consequently, the choice of SA topology significantly impacts the footprint of the memory array. Considering that SRAM density is governed by the employed SA, careful consideration of area becomes a critical aspect when selecting the appropriate sensing scheme. The layout of MDSA and MDSANF is specifically designed to estimate the area requirements of the corresponding SA as illustrated in figure 11a and

Just Accepted

Table 1. Performance Summary of Conventional SA designs with MDSA and MDSANF at 0.8 V supply voltage

SA Design	Read Delay (ps)	Power Consumption (μW)	PDP (fJ)	Normalized Area
CBLSA	1105.92	6.33	9.82	1.4
HCSA	761.03	7.80	5.93	2.55
MCSA	561.84	2.23	1.25	1.3
MHSA	644.23	30.81	19.85	3.3
LSA	580.81	1.90	1.10	1
CLSA	1575.29	15.57	24.53	1.16
DLSA	706.55	16.16	11.41	1.65
DTSA	689.15	1.70	1.17	2.225
MDSA	**401.13**	**5.50**	**2.21**	**1.25**
MDSANF	**415.10**	**5.80**	**2.49**	**1.5**



(a) (b)

Fig. 11. Layout design for area estimation of (a) MDSA and (b) MDSANF

11b respectively. The estimated area for MDSA is 0.297 μm^2 whearas MDSANF occupies 0.5192 μm^2. The area increase for MDSANF is due to the larger sized NMOS foot-switch.The performance summary for MDSA and MDSANF together with conventional SA designs is given in Table 1. The estimated area for all SAs is normalized to that of LSA, which is the earliest iterations of SA built around the inverter-latch. The MHSA has largest area due to the large sized PMOS used to build the current conveyor whereas the dual tail sensing scheme used in DTSA is a major contributor to its larger area footprint.

5 CONCLUSION

In this work, a pair of fast sensing modified decoupled sense amplifiers are proposed. The impact of various parameter variations on the SRAM read performance and SA reliability is investigated using Monte Carlo simulation for 10k data points. The delay and power consumption together with leakage power are calculated for conventional SAs at 22 nm technology node. The bit line capacitance and differential bit line voltage variation are considered together with global process corners to justify the robustness of the SA architectures. Threshold voltage variation causes transistor mismatch and can lead to read failures. The optimal sizing of decoupling PMOS has also been highlighted and its effect on other performance parameters such as sensing delay and power consumption is calculated. The MDSA and MDSANF show read delay improvement of 37.74% and 35.58% over

Just Accepted

MHSA at V_{DD} = 0.8 V and bit line capacitance of 50 fF. The MDSA and MDSANF thus provide a high speed read operation at supply voltages ranging from 0.5 to 1 when compared to conventional SAs.

REFERENCES

Erfan Abbasian and Morteza Gholipour. 2022. A low-leakage single-bitline 9T SRAM cell with read-disturbance removal and high writability for low-power biomedical applications. *International Journal of Circuit Theory and Applications* 50, 5 (2022), 1537–1556.

Nanoscale Integration Arizona State University and Modeling Group. 2008. Predictive Technology Models. http://ptm.asu.edu/latest.html

Azeez J. Bhavnagarwala, Xinghai Tang, and James D. Meindl. 2001. The impact of intrinsic device fluctuations on CMOS SRAM cell stability. *IEEE Journal of Solid-State Circuits* 36 (2001). Issue 4.

Travis N Blalock and Richard C Jaeger. 1991. A high-speed clamped bit-line current-mode sense amplifier. *IEEE Journal of Solid-State Circuits* 26, 4 (1991), 542–548.

PY Chee, PC Liu, and L Siek. 1992. High-speed hybrid current-mode sense amplifier for CMOS SRAMs. *Electronics Letters* 9, 28 (1992), 871–873.

Jian Chen, Wenfeng Zhao, Yuqi Wang, and Yajun Ha. 2021. Analysis and optimization strategies toward reliable and high-speed 6T compute SRAM. *IEEE Transactions on Circuits and Systems I: Regular Papers* 68, 4 (2021), 1520–1531.

AT Do, ZH Kong, and KS Yeo. 2007. 0.9 V current-mode sense amplifier using concurrent bit-and data-line tracking and sensing techniques. *Electronics Letters* 43, 25 (2007), 1421–1422.

Divya Dutt, Poornima Mittal, Bhawna Rawat, and Brijesh Kumar. 2022. Design and Performance Analysis of High-Performance Low Power Voltage Mode Sense Amplifier for Static RAM. *Advances in Electrical and Electronic Engineering* 20, 3 (2022), 285–293.

Roman Fragasse, Ramy Tantawy, Brian Dupaix, Trevor Dean, Daron Disabato, Matthew R. Belz, Dale Smith, Jamin McCue, and Waleed Khalil. 2019. Analysis of SRAM enhancements through sense amplifier capacitive offset correction and replica self-timing. *IEEE Transactions on Circuits and Systems I: Regular Papers* 66 (2019). Issue 6. https://doi.org/10.1109/TCSI.2019.2902102

Carles Hernández, Antoni Roca, Jose Flich, Federico Silla, and Jose Duato. 2011. Characterizing the impact of process variation on 45 nm NoC-based CMPs. *J. Parallel and Distrib. Comput.* 71, 5 (2011), 651–663.

Muhammad M Khellah and Mohamed I Elmasry. 2001. A low-power high-performance current-mode multiport SRAM. *IEEE transactions on very large scale integration (VLSI) systems* 9, 5 (2001), 590–598.

Tsuguo Kobayashi, Kazutaka Nogami, Tsukasa Shirotori, and Yukihiro Fujimoto. 1993. A current-controlled latch sense amplifier and a static power-saving input buffer for low-power architecture. *IEICE transactions on electronics* 76, 5 (1993), 863–867.

Rohit Lorenzo and Roy Paily. 2022. Half-selection disturbance free 8T low leakage SRAM cell. *International Journal of Circuit Theory and Applications* 50, 5 (2022), 1557–1575.

Poornima Mittal, Bhawna Rawat, and Nishant Kumar. 2022. Tetra-variate scrutiny of diverse multiplexer techniques for designing a barrel shifter for low power digital circuits. *Microprocessors and Microsystems* 90 (2022), 104491.

Dhruv Patel, Adam Neale, Derek Wright, and Manoj Sachdev. 2021. Body Biased Sense Amplifier With Auto-Offset Mitigation for Low-Voltage SRAMs. *IEEE Transactions on Circuits and Systems I: Regular Papers* 68, 8 (2021), 3265–3278.

Bhawna Rawat and Poornima Mittal. 2021. Single bit line accessed high-performance ultra-low voltage operating 7T static random access memory cell with improved read stability. *International Journal of Circuit Theory and Applications* 49, 5 (2021), 1435–1449.

Bhawna Rawat and Poornima Mittal. 2022a. A comprehensive analysis of different 7T SRAM topologies to design a 1R1W bit interleaving enabled and half select free cell for 32 nm technology node. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 478, 2259 (2022), 20210745.

Bhawna Rawat and Poornima Mittal. 2022b. A reliable and temperature variation tolerant 7T SRAM cell with single bitline configuration for low voltage application. *Circuits, Systems, and Signal Processing* 41, 5 (2022), 2779–2801.

Bhawna Rawat and Poornima Mittal. 2022c. A reliable and temperature variation tolerant 7T SRAM cell with single bitline configuration for low voltage application. *Circuits, Systems, and Signal Processing* 41, 5 (2022), 2779–2801.

Bhawna Rawat and Poornima Mittal. 2022d. A Switching NMOS Based Single Ended Sense Amplifier for High Density SRAM Applications. *ACM Trans. Des. Autom. Electron. Syst.* (dec 2022). Just Accepted.

Bhupendra Singh Reniwal, Vikas Vijayvargiya, Pooran Singh, Nand Kishor Yadav, Santosh Kumar Vishvakarma, and Devesh Dwivedi. 2019. An Auto-Calibrated Sense Amplifier with Offset Prediction Approach for Energy-Efficient SRAM. *Circuits, Systems, and Signal Processing* 38 (2019). https://doi.org/10.1007/s00034-018-0934-1

Hans Reyserhove and Wim Dehaene. 2017. A Differential Transmission Gate Design Flow for Minimum Energy Sub-10-pJ/Cycle ARM Cortex-M0 MCUs. *IEEE Journal of Solid-State Circuits* 52, 7 (2017), 1904–1914. https://doi.org/10.1109/JSSC.2017.2693241

Daniel Schinkel, Eisse Mensink, Eric Klumperink, Ed Van Tuijl, and Bram Nauta. 2007. A double-tail latch-type voltage sense amplifier with 18ps setup+ hold time. In *2007 IEEE international solid-state circuits conference. Digest of technical papers*. IEEE, 314–605.

Evert Seevinck, Petrus J van Beers, and Hans Ontrop. 1991. Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAM's. *IEEE Journal of Solid-State Circuits* 26, 4 (1991), 525–536.

Just Accepted

Jaspal Singh Shah, David Nairn, and Manoj Sachdev. 2013. An Energy-Efficient Offset-Cancelling Sense Amplifier. *IEEE Transactions on Circuits and Systems II: Express Briefs* 60, 8 (2013), 477–481. https://doi.org/10.1109/TCSII.2013.2268312

Vibhu Sharma, Stefan Cosemans, Maryam Ashouei, Jos Huisken, Francky Catthoor, and Wim Dehaene. 2011. A 4.4 pJ/access 80 MHz, 128 kbit variability resilient SRAM with multi-sized sense amplifier redundancy. *IEEE Journal of Solid-State Circuits* 46, 10 (2011), 2416–2430.

Shan Shen, Hao Xu, Yongliang Zhou, and Wenjian Yu. 2022. A Single-Ended Offset-Canceling Sense Amplifier Enabling Wide-Voltage Operations. *IEEE Transactions on Circuits and Systems II: Express Briefs* (2022), 1. https://doi.org/10.1109/TCSII.2022.3219136

Naveen Verma and Anantha P Chandrakasan. 2007. A 65nm 8T sub-Vt SRAM employing sense-amplifier redundancy. In *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*. IEEE, 328–606.

Jinn-Shyan Wang and Hong-Yu Lee. 1998. A new current-mode sense amplifier for low-voltage low-power SRAM. In *Proceedings Eleventh Annual IEEE International ASIC Conference (Cat. No. 98TH8372)*. IEEE, 163–167.

Bernhard Wicht, Thomas Nirschl, and Doris Schmitt-Landsiedel. 2004. Yield and speed optimization of a latch-type voltage sense amplifier. *IEEE Journal of Solid-State Circuits* 39, 7 (2004), 1148–1158.

Yue Zhao, Jinkai Wang, Zhongzhen Tong, Xiulong Wu, Chunyu Peng, Wenjuan Lu, Qiang Zhao, and Zhiting Lin. 2022. An offset cancellation technique for SRAM sense amplifier based on relation of the delay and offset. *Microelectronics Journal* 128 (10 2022), 105578.

Hong Zhu and Volkan Kursun. 2014. A Comprehensive Comparison of Data Stability Enhancement Techniques With Novel Nanoscale SRAM Cells Under Parameter Fluctuations. *IEEE Transactions on Circuits and Systems I: Regular Papers* 61, 5 (2014), 1473–1484. https://doi.org/10.1109/TCSI.2013.2289411

Jiafeng Zhu, Na Bai, and Jianhui Wu. 2013. A review of sense amplifiers for static random access memory. *IETE Technical Review* 30, 1 (2013), 72–81.

Received 20 January 2023

Just Accepted

Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023)
IEEE Xplore Part Number: CFP23AWO-ART; ISBN: 979-8-3503-9663-8

Multilabel News Category Classification using Machine Learning

Roshan Kumar Shah
Dept. of CSE
Delhi Technological University
Delhi, India-110042
i.roshankrshah@gmail.com

Sahil Kumar
Dept. of CSE
Delhi Technological University
Delhi, India-110042
sahilkumar6102001@gmail.com

Shashank
Dept. of CSE
Delhi Technological University
Delhi, India-110042
shashankkr550@gmail.com

2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192826

***Abstract*—Multi-label text classification is a technique used to categorize texts into multiple labels or categories. Such classification problems are difficult as well as captivated with several parameters and changes per the matter. News categories classification is one of the most popular and necessary multi-label text classification problems because the same news might or might not belong to multiple genres at constant times. The majority of the factors of news genre classification are based on parameters like news headlines, articles, and their short descriptions. The project analyzes news articles using a variety of machine learning and natural language processing approaches. The goal is to take a big collection of article headlines, group them into similar, related stories, and then analyze those groups further to try to develop a statistic for how well the collection of articles covers various points of view on the issue. This uses a variety of Natural Language Processing techniques to pre-process the text in order to improve the amount of importance that can be deduced. It is implemented in Python. It makes use of freely accessible NLP libraries like NLTK.**

Keywords: News Category Classification, Multilabel Text Classification, Label Powerset, Machine Learning Classifiers, Binary Relevance

I. Introduction

In recent years, natural language processing (NLP) has become a very popular research domain and provides numerous real-life problems like fake news detection, rumor prediction, sentiment analysis, multi-label text classification (MLTC), and many others [1], [2]. Multi-label text classification (MLTC) techniques are widely employed nowadays in a variety of real-world applications, such as recommender systems, music categorization, movie genre prediction, and semantic scene classification [3]. The most common NLP task, known as "multi-label text categorization," aims to assign numerous labels to each piece of input text. In recent years, Machine learning and deep learning techniques have been pivotal in solving many real-life problems like medical imaging, social network analysis, text classification, text summarization, and others [4]–[7].

The paper analyzes news articles using a variety of machine learning and natural language processing approaches. The goal is to take a big collection of article headlines, group them into similar, related stories, and then analyze those groups further to try to develop a statistic for how well the collection of articles covers various points of view on the issue. In recent times, primarily several neural network-based approaches are planned for the determination of such issues however they're resource-intensive and time intense. Here we have a tendency to try and propose a unique methodology of news category classification. The overall work introduced in the paper can be divided into two main phases: Use topic mining or clustering to examine news articles and organize them into inferred-related stories. This uses a variety of Natural Language Processing techniques to pre-process the text in order to improve the amount of importance that can be deduced. It is implemented in Python. It makes use of freely accessible NLP libraries like NLTK. For the greatest outcomes, the intent was to apply the most recent machine learning and natural language processing methods.

II. Problem Statement

A. Background

Multi-label classification is the conception of a single-label problem, and a single instance can belong to more than one single class. From the official documentation of the Scikit-learn package, "This can be allowed of as predicting parcels of a sample that aren't mutually exclusive." [8] There are no constraints on the number of classes that a case can be assigned to in a multi-label classification problem. In an analogous environment, there exists the multi-class bracket problem. Still, the crucial difference between both is the fact that the multi-label bracket supposes that the parcels aren't mutually exclusive.

B. Problem transformation Technique

The multi-label dataset is changed into a single-label dataset in this phrase. Machines can easily read single-label datasets and issues, which facilitates model creation. The following methods are used to transform problems:

1) Binary Relevance (BR): This method treats each label separately and divides the multi-label data into various single-class classification tasks. The multi-label problem is split into three separate single-class classification problems in order to be addressed by binary relevance. [9] A backup plan is offered in case the first technique fails due to label correlation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:33:22 UTC from IEEE Xplore. Restrictions apply.

2) Label Powerset: Label Powerset is a very simple and less used approach in Problem transformation but it has very useful advantages in this approach there's a multilabel problem into a single multi-class problem with 2L labels and class values. After the conversion into a multi-class problem, it can be classified with any multiclass classifier technique. [9] In this approach, there is no information loss and also it considers the label dependency and correlation btw data and the labels

3) Classifier chain: It is one of the techniques used in problem transformation in multi-label classification. It retains the ability to account for label dependencies for classification while combining the computational effectiveness of the Binary Relevance technique. Based on the BR approach, the Classifier Chains method is effective even with a large number of labels. It also takes into account how different labels are related to one another. [9]

III. Text Vectorizers

A common technique for transforming input data from its original text-based form into real-number vectors is called vectorization. This format is supported by Machine Learning models. It's one of the most efficient techniques for the respective purpose.

A. Count Vectorization

This technique creates a matrix having a column size equal to the unique word in the whole given document and a row size equal to the no. of the different text samples in that given document. Each cell in the table represents the count of words that are present in the corresponding text sample.

B. TF-IDF

It is based on the frequency (count) of a word in the corpus but it also provides importance of word with respect of other words in term of numeric value i.e, higher the importance higher the numeric value or lower the importance lower the numeric value. Term frequency adjusted for document length is determined by dividing the total number of words in the document by the raw count of occurrences. [10]

TF: It keeps track of how frequently each word appears in each document. The frequency of occurrence of the phrase t in document d is represented by tf(t,d). Phrase t appears numerous times in document d. [11]

$$tf(t,d) = \frac{f_d(t)}{\max_{w \in d} f_d(w)} \tag{1}$$

Equation 1: Term Frequency

IDF: Depending on the word's frequency, which reveals how common or unusual it is, measures how much information is being sent by word processing.. [11]

$$idf(t,D) = \ln\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right) \tag{2}$$

Equation 2: Inverse Document Frequency

TF-IDF is calculated as: [4]

$$tfidf(t,d,D) = tf(t,d) \cdot idf(t,D) \tag{3}$$

Equation 3: Term frequency-inverse document frequency

IV. Machine Learning Classifiers

A. Logistic Regression:

It is a well-known machine learning algorithm. The value of a categorical dependent variable is predicted using a set of independent variables. Based on the categorical dependent value, it makes predictions. It outputs probabilistic values, such as 0 or 1, as output.

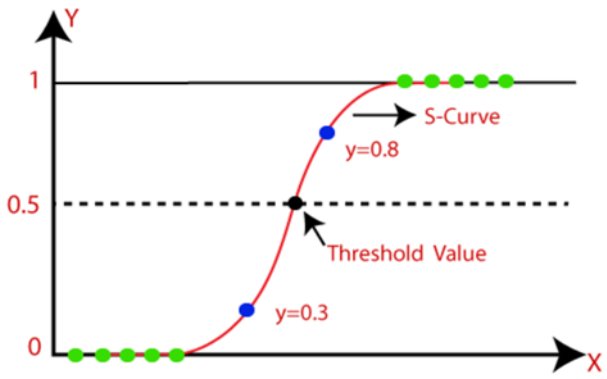


Fig. 1. Graphical Representation of Logistic Regression [13]

B. Multinomial Naïve Bayes

In the field of Natural Language Processing, the Multinomial Naive Bayes algorithm is a well-known method for Bayesian learning. It makes predictions about the categorization of textual data, such as news articles or emails, using the concepts of the Bayes theorem.

By calculating the likelihood of each potential tag for a specific sample, the algorithm selects the tag with the highest probability as the output.

This method has gained popularity due to its effectiveness in text classification tasks. [14] The Bayes theorem is used as the only basis for the multinomial model's evaluation of probability-based decision functions. The Bayesian probability is computed as follows:

$$p(C_k|W) = \frac{p(C_k) \cdot p(W|C_k)}{p(W)} \tag{4}$$

Equation 4: Bayesian Probability

C. Linear SVC

The Linear Support Vector Classifier (SVC) is a classification method that uses a linear kernel function to make predictions. The main advantage of this is that it also performs well with large data samples. In essence, Linear Support Vector Machine (Linear SVC) is an algorithm that

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:33:22 UTC from IEEE Xplore. Restrictions apply.

looks for a hyperplane to maximize the distance between samples that are classified. [15]

V. PROPOSED METHODOLOGY

A. About Dataset

A pretty diverse dataset was used that was collected from a variety of websites, including Kaggle [16], India Today [17], the BBC [18], Google News [19] etc. Various categories of news related were commbined into a single dataset after gathering them from their separate sources. Adding a little more detail to this, the issue is that there isn't a single dataset that includes all the different news categories that was intended to include in this project. We primarily sought to collect the following 3 attributes from this dataset: "Headlines," "Description," and "Categories." Hence, all of those datasets were merged to analyze on a very wide spectrum of news categories and normalize the categories to their supercategories. After this, 24 news categories were labeled. There are three datasets from Google News, in short News and CNN which contains approximately 60,000 articles. One dataset was used for training purposes and the other two datasets were used for testing the models.

ARTS & CULTURE	EDUCATION	LIFESTYLE	STYLE & BEAUTY
BUSINESS & FINANCE	ENTERTAINMENT	MEDIA	TECHNOLOGY
'COLLEGE	ENVIRONMENT	POLITICS	TOURISM
COMEDY	FOOD & BEVERAGES	RELIGION	WEDDINGS
CRIME	HEALTH & FITNESS	SCIENCE	WELLNESS
CURRENT AFFAIRS	HEALTHY LIVING	SPORTS	WOMEN

Fig. 2. News Categories in the Dataset

B. DATA PREPROCESSING

The data needed to be pre-processed so that it could be changed into forms that the algorithm could identify after all of the data had been integrated.

1. Verify that no fields with missing values exist in the dataset being used.
2. Encoding of Category: An encoder called "Multiple label Binarizer" was employed to transform the category column of the dataset which was in set of string format into 24 category wise columns which contains binary value where 1 in these columns indicates the belonging category whereas 0 indicates not belonging category of the news articles.
3. Feature engineering: The headline and article columns were combined because they both included the same information in a single column of text, and further operations were then made to the text column.
4. Data cleaning:

- Html tags removal Tag = <.*?>
- Removing special characters and digits
- Using the NLTK package, remove English stop words and punctuation.
- Lemmatizing words: To get to a word's fundamental form, word affixes were deleted. All words should be lowercase.

After these steps the various combination of text vectorizers and machine learning algorithms were applied on the preprocessed dataset.

VI. TRAINING MODELS:

Splitting the dataset into training and validation halves was therefore the first and foremost task. The problem transformation approach was then used (PTA). Three PTA methods—Binary Relevance, Label PowerSet, and Classifier Chain—have been employed. Different combinations of text vectorizers and ML classifiers have been untilized, including binary classifiers, for the implementation of these three. Due to the model's use of two different vectorizers (Count Vectorization and TF-IDF) and three different classifiers (Logistic Regression, MultinomialNB, and Linear SVC), 15 combinations in all are created for validation, testing, and evaluation. The three PTA methods that have been used—binary relevance, label powerset, and classifier chain—are briefly described here.

A. Binary Relevance

With BR two vectorization algorithms named CV and TF IDF have been utilized each time. Then, utilizing several machine learning classifiers such as LR, MNB, and SVC, each were combined, such as BR- CV and BR- TF IDF. Then a contrast was made between the F1- score, precession, and accuracy.

B. Label PowerSet (LP)

With LP two vectorization methods named CV and TF-IDF were used each time. Then with the combination of each like LP- CV, LP- TF-IDF, various machine learning classifiers named LR, MNB, SVC, were used and using these combinations. Then a comparison was made between the accuracy, precession, and F1- score.

C. Classifier Chain

With BR, the vectorization techniques Count Vectorizer and TF IDF were employed alternatively. Then, utilizing several machine learning classifiers such as LR, MNB, and SVC, each were combined, such as BR- CV and BR- TF IDF. then a contrast was made between the F1- score, precession, and accuracy.

VII. RESULT AND ANALYSIS

A. Result & accuracy

The evaluation of multi-label text categorization techniques has been approached in a number of different ways. Utilizing just one evaluation metric can lead to inaccurate interpretations of the metrics by providing false insights into the outcome. The

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:33:22 UTC from IEEE Xplore. Restrictions apply.

parameters that are often measured were also averaged, such as Accuracy, Precision, Recall, and F1-score, by averaging the metrics over all genres.

Accuracy$_{avg}$: "The average accuracy for each instance is calculated as the ratio of successfully predicted output to all other possible outputs." [20]

$$Accuracy_{avg} = \frac{\sum_{i=1}^{n}(A_i \times K_i)}{\sum_{i=1}^{n} K_i} \quad (5)$$

Equation 5: Average Accuracy

Precision$_{avg}$: "It represents the average across all instances of the ratio of projected accurate labels to the total number of actual labels." [20]

$$Precision_{avg} = \frac{\sum_{i=1}^{n}(P_i \times K_i)}{\sum_{i=1}^{n} K_i} \quad (6)$$

Equation 6: Average Precision

Recall$_{avg}$: "The ratio of accurate labels to all predicted labels, on average across all cases, is known as a recall" [20]

$$Recall_{avg} = \frac{\sum_{i=1}^{n}(R_i \times K_i)}{\sum_{i=1}^{n} K_i} \quad (7)$$

Equation 7: Average Recall

F1-score$_{avg}$: "It is the harmonic mean of Precision and Recall " [20]

$$F1score_{avg} = \frac{\sum_{i=1}^{n}(F1_i \times K_i)}{\sum_{i=1}^{n} K_i} \quad (8)$$

Equation 8: Average F1-score

The experiment was carried out and the value of several assessment criteria was measured using the test dataset. On the numerous datasets under consideration, multiple news genre classification models were applied. The experiment is run on a laptop running the Windows 11 Operating System, an Intel i5-7400H processor with four 3.0GHz CPUs, 8GB of RAM, and 512GB of SSD storage. Scikit-learn, Numpy, Pandas, and NLTK are some of the packages which were used to run the simulations in Python.

B. Result based on Binary Relevance method

The result was achieved using the Binary Relevance based technique in conjunction with the "CV and TF-IDF", as well as the ML algorithms LR, MNB, and LinearSVC.In both datasets, the BR, CV, and Linear SVC model combination had the highest accuracy, precision, and F1-score, which was approximately 0.72, 0.84, and 0.86.

The graphical representation of the results of both datasets is shown figure 3,4 and 5 using Binary Relevance i.e Accuracy, F1-Score, and Precision.

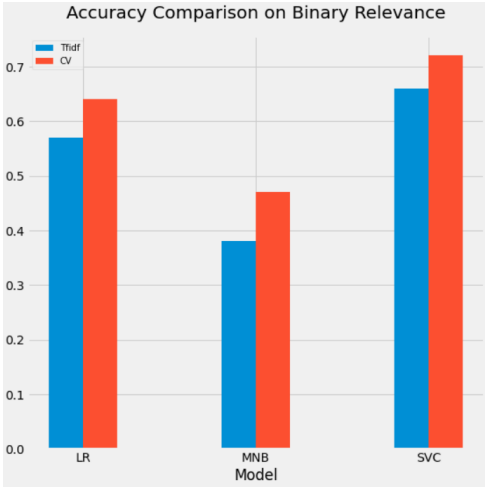


Fig. 3. Accuracy comparison on Binary Relevance on Google News Dataset

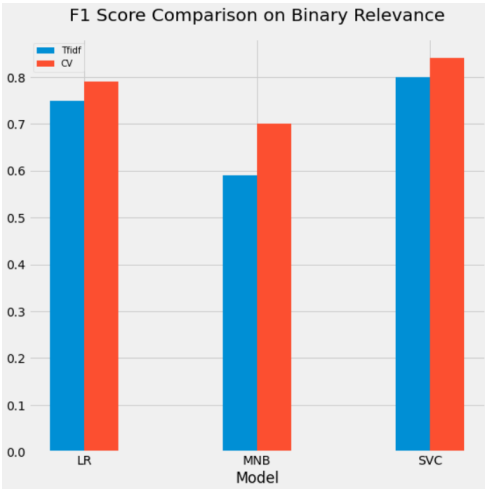


Fig. 4. F1-Score comparison on Binary Relevance on Google News Dataset

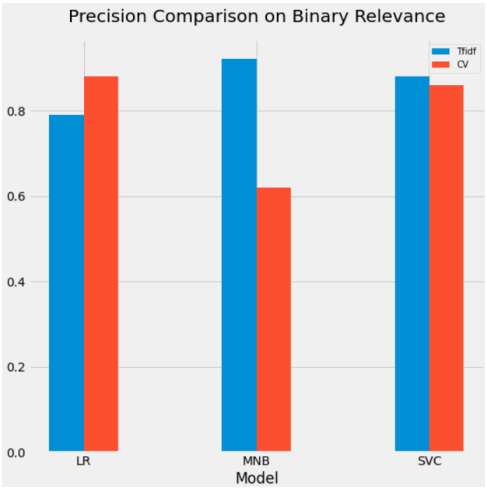


Fig. 5. Precision comparison on Binary Relevance on Google News Dataset

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:33:22 UTC from IEEE Xplore. Restrictions apply.

C. Result based on Label Power set method

The result was achieved using the Label PowerSet-based technique in conjunction with the "CV and TF-IDF",as well as the ML algorithms LR, MNB, and LinearSVC. In both datasets, the LP, TFidf, and Linear SVC model combination had the highest accuracy, precision, and F1- score, which was approximately 0.91, 0.92 and 0.91.
The graphical representation of the results of both datasets is shown figure 6, 7 and 8 using Label Powerset i.e Accuracy, F1-Score, and Precision.

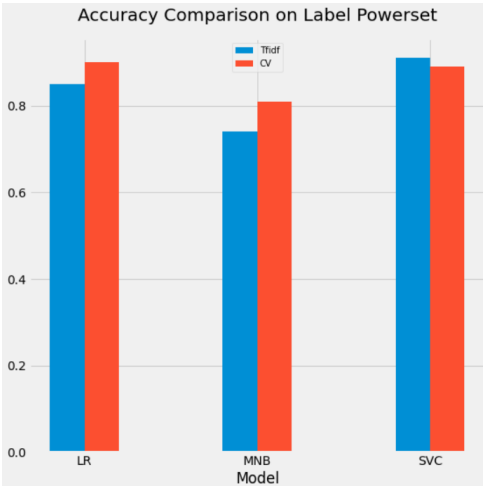


Fig. 6. Accuracy comparison on Label PowerSet Google News Dataset

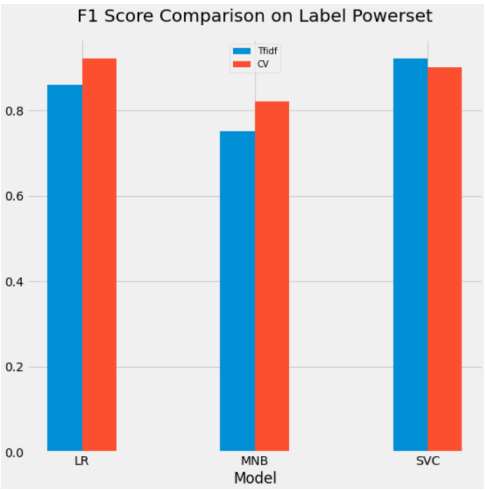


Fig. 7. F1-Score comparison on Label PowerSet Google News Dataset

D. Result based on Classifier chain method

Because this method was more sophisticated and took longer to provide results, it used fewer combinations than the method described above. In the prior method, however, combinations with higher precision were used i.e, "CV and TF-IDF", with LR and LinearSVC algorithm. In both datasets, the CC, Tf-idf, and Linear SVC model combination had the highest accuracy, precision, and F1- score, which was approximately 0.25, 0.27 and 0.28.

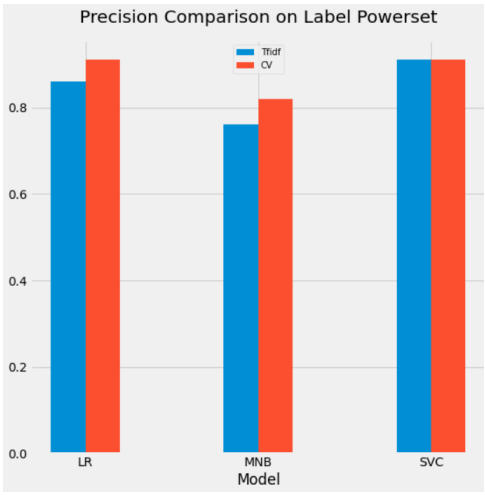


Fig. 8. Precision comparison on Label PowerSet Google News Dataset

E. Result Comparison

The results for accuracy, precision, and F1-score using 15 combinations of various approaches are shown in Tables 1 and 2. By comparing the values of this method the most efficient model that was obtained of the combination of LP, TFidf, and Linear SVC having accuracy, precision, and F1-score on Google News dataset 0.91, 0.92, 0.91 respectively and on Inshort News dataset 0.91, 0.92, 0.91. on both datasets, approximately the same result was achieved. The second highest efficient model was a combination of LP, CV, and Linear SVC having accuracy, precision, and F1-score on Google News datasets 0.90, 0.91, and 0.91, respectively. And On the Inshort News dataset 0.90, 0.91, 0.91 respectively on both datasets approximately the same result was achieved.

TABLE I
RESULTS USING GOOGLE NEWS DATASET.
PTA: PROBLEM TRANSFORMATION APPROACH. CV: COUNTVECTORIZER. TF-IDF: TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY. LR: LOGISTIC REGRESSION. MNB: MULTINOMIAL NAIVE BAYES. SVC: SUPPORT VECTOR CLASSIFICATION. CC: CLASSIFIER CHAIN.

S.no	PTA	Accuracy	F1 Score	Precision
1	LP, CV,LR	0.90	0.92	0.91
2	LP, Tf-idf, LR	0.85	0.86	0.86
3	LP, CV, MNB	0.81	0.82	0.82
4	LP, Tfidf, MNB	0.74	0.75	0.76
5	LP, CV, linear svc	0.89	0.90	0.91
6	**LP, Tfidf, linear svc**	**0.91**	**0.92**	**0.91**
7	BR, Tf-idf, LR	0.57	0.75	0.79
8	BR, CV, LR	0.64	0.79	0.89
9	BR, Tf-idf, MNB	0.38	0.59	0.92
10	BR, CV, MNB	0.47	0.70	0.62
11	BR, Tf-idf, Linear svc	0.66	0.80	0.88
12	BR, CV, Linear svc	0.72	0.84	0.86
13	CC, Tf-idf,LR	0.23	0.25	0.24
14	CC, Tf-idf,SVC	0.25	0.27	0.28
15	CC,CV,SVC	0.25	0.27	0.28

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:33:22 UTC from IEEE Xplore. Restrictions apply.

TABLE II
RESULTS USING IN SHORT NEWS DATASET.
PTA: PROBLEM TRANSFORMATION APPROACH. CV: COUNTVECTORIZER. TF-IDF: TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY. LR: LOGISTIC REGRESSION. MNB: MULTINOMIAL NAIVE BAYES. SVC: SUPPORT VECTOR CLASSIFICATION. CC: CLASSIFIER CHAIN.

S.no	PTA	Accuracy	F1 Score	Precision
1	LP, CV, LR	0.91	0.91	0.90
2	LP, Tf-idf, LR	0.86	0.86	0.85
3	LP, CV, MNB	0.81	0.83	0.82
4	LP, Tfidf, MNB	0.74	0.76	0.74
5	LP, CV, linear svc	0.89	0.90	0.91
6	**LP, Tfidf, linear svc**	**0.91**	**0.92**	**0.90**
7	BR, Tf-idf, LR	0.57	0.75	0.94
8	BR, CV, LR	0.64	0.79	0.88
9	BR, Tf-idf, MNB	0.39	0.60	0.92
10	BR, CV, MNB	0.47	0.71	0.62
11	BR, Tf-idf, Linear svc	0.66	0.80	0.88
12	BR, CV, Linear svc	0.71	0.82	0.86
13	CC, Tf-idf,LR	0.20	0.20	0.22
14	CC, Tf-idf,SVC	0.20	0.20	0.22
15	CC,CV,SVC	0.20	0.20	0.22

VIII. CONCLUSION

In this paper, multilabel news classification was introduced by utilizing three PTAs called "Binary Relevance, Label Powerset Classifier Chain," two TVs called "count vectorizer and TF-IDF," and three machine learning (ML) models for classification called "logistic regression (LR). In essence, several PTA were combined, "text vectorizers, and machine learning classifiers," and have come up with a total of 15 unique combinations that categorize news into the right genres. In summary, it may be stated that the optimum result was obtained by using the "support vector classifier (SVC) as a machine learning model, label powerset (LP) as a problem transformation technique, and TF-IDF as a text vectorizer technique." Hence, it produced an accuracy of about 91%, an F1 score of about 92%, and an overall precision of about 90% on the news datasets.

REFERENCES

[1] S. Kumar, A. Kumar, A. Mallik, R. R. Singh, "OptNet-Fake: Fake News Detection in Socio-Cyber Platforms Using Grasshopper Optimization and Deep Neural Network". IEEE Transactions on Computational Social Systems, 2023
[2] A. Kumar, N. Aggarwal, S. Kumar "SIRA: a model for propagation and rumor control with epidemic spreading and immunization for healthcare 5.0". Soft Computing, 27, 4307–4320, 2023
[3] S. Kumar, N. Kumar, A. Dev, S. Naorem (2023). "Movie genre classification using binary relevance, label powerset, and machine learning classifiers". Multimedia Tools and Applications, 82(1), 945-968, 2023
[4] S. Kumar, A. Mallik, B. S. Panda, "Influence maximization in social networks using transfer learning via graph-based LSTM". Expert Systems with Applications, 212, 118770, 2023
[5] S. Anand, A. Mallik, S. Kumar, "Integrating node centralities, similarity measures, and machine learning classifiers for link prediction". Multimedia Tools and Applications, 81(27), 38593-38621, 2022
[6] S. Kumar, A. Mallik, "COVID-19 Detection from Chest X-rays Using Trained Output Based Transfer Learning Approach". Neural processing letters, 1-24, 2022
[7] S. Kumar, A. Kumar, A. Mallik, S. Dhall, "Opinion leader detection in Asian social networks using modified spider monkey optimization". Transactions on Asian and Low-Resource Language Information Processing, 2022
[8] S.Li, "Towards Data Science,"2018. Available: https://towardsdatascience.com/multi-label-text-classification-with-scikit-learn-30714b7819c5.
[9] C.Kariuki,2021. https://www.section.io/engineering-education/multi-label-classification-with-scikit-multilearn/.
[10] A. Simha, 2021. [Online]. Available: https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/.
[11] D.Tomruk, "Towards Data Science," 2020. [Online]. Available: https://towardsdatascience.com/tf-idf-term-frequency-and-inverse-dense-frequency-techniques-472bf1ba311b.
[12] P. Jain, 2021. [Online]. Available: https://towardsdatascience.com/basics-of-countvectorizer-e26677900f9c.
[13] "Javapoint," [Online]. Available: https://www.javatpoint.com/logistic-regression-in-machine-learning.
[14] A.V.Ratz,2021. [Online]. Available:https://towardsdatascience.com/multi-nomial
[15] M. S. A. a. M. M. F. Seyyed Mohammad Hossein Dadgar, "A novel text mining approach based on TF-IDF and Support Vector Machine for news classification," IEEE, 2016.
[16] P.SHARIF, "Kaggle," 2017. https://www.kaggle.com/datasets/pariza/bbc-news-summary.
[17] "India Today Group," [Online]. Available: https://www.indiatodaygroup.com/.
[18] "BBC India," [Online]. Available: https://www.bbc.com/news.
[19] "Google News," Google, [Online]. Available: https://news.google.co.in/.
[20] R.Vickery, "Towards Data Science," 2021. [Online]. Available: https://towardsdatascience.com/8-metrics-to-measure-classification-performance-984d9d7fd7aa.
[21] Y. Y. T. Li, "Word Embedding for Understanding Natural Language," Springer, Cham, 2018.
[22] B. F. Y. E. N. C. Sahar Ghannay, "Word Embedding Evaluation and Combination," European Language Resources Association (ELRA), 2016.
[23] P. B. M. TÜFEKCİ, "Author and genre identification of Turkish news texts using deep learning algorithms.," Springer, 2022.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:33:22 UTC from IEEE Xplore. Restrictions apply.

Multimodal Sarcasm Detection (MSD) in Videos using Deep Learning Models

Ananya Pandey
Department of Information Technology,
Delhi Technological University
Bawana Road, New Delhi-110042, India
ananyaphdit08@gmail.com

Dinesh Kumar Vishwakarma
Department of Information Technology,
Delhi Technological University
Bawana Road, New Delhi-110042, India
dvishwakarma@gmail.com

2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT) | 979-8-3503-3936-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/APSIT58554.2023.10201731

***Abstract*— Every day, individuals all over the globe use video-sharing websites to broadcast their thoughts, experiences, and recommendations to the world. There has been a huge rise in educational interest in the area of sarcasm detection for these opinionated videos. Although sarcasm has proven effective for written text, it remains an unexplored area of research when applied to video and other forms of multimodal data. Numerous verbal and nonverbal signs, such as a shift in voice, an overemphasis in a phrase, a stretched pronunciation, or a stiff-looking face, are often used to convey sarcasm. The majority of current research on sarcasm detection has heavily relied on textual content. Hence, in this paper, we suggest that the use of multiple modalities can help with more accurate sarcasm identification. The results of our studies on one of the multimodal sarcasm detection (MSD) benchmark datasets "MUSTARD" reveal that our strategy outperforms other algorithms in sarcasm classification.**

Keywords—Multimodal Sarcasm Detection (MSD), Multimodal, RegNet, DistilBERT, BERT

I. INTRODUCTION

Multimodal Sarcasm Detection (MSD) is a problem that involves predicting class labels as non-sarcastic or sarcastic based on three modalities: the textual, the visual, and the audio content associated with the videos. The ability to ridicule or show hate in regular communication is facilitated by the use of sarcasm. This is accomplished by using ironic phrases which represents an adverse perception. A sarcastic comment like: "I work 40 hours a week for me to be this impoverished", represents nothing but sarcasm. Nonetheless, there are times when sarcasm lacks unambiguous linguistic cues, Hence, there is a need for other clues to understand a speaker's intentions more accurately. For example, sarcasm may be portrayed through a variety of vocal and non-vocal signs, for instance a shift in voice, an overemphasis in a phrase, a stretched pronunciation, or a stiff-looking face. Besides this, sarcasm detection includes determining verbal or contextual incongruity, from multiple modalities [1],[2].

The feature-level fusion technique is used to combine the multimodal data retrieved by [3]. To understand how the various parts of an utterance relate to one another regardless of modality, [4] introduced an attention model. The incompatibility across modalities, which each express a different mood, is a key indication for detecting sarcasm; yet, these works do not take into account its significance. The cross-domian sarcastic utterance shown in **Figure 1** helps us communicate our logic more clearly. It's easy to see from the audio signal that the speaker is emphasizing the word "incredible". The speaker, meanwhile, is raising his eyes. Hence, sarcasm is often shown by the mismatch between the text (a positive term) and the visual (an eye raise) or audio signal. Hence, sarcasm detection can be improved by focusing on multimodal aspects.

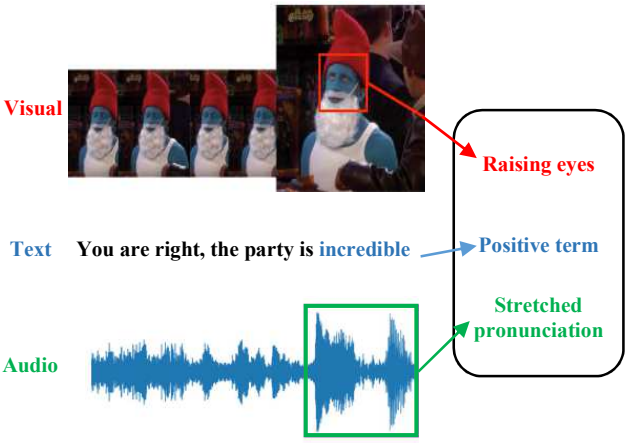


Figure 1 Sarcastic example using multiple modalities

Lately, there has been a rise in interest in MSD, which seeks to characterize a video as sarcastic or non-sarcastic by taking into account: textual, visual, and acoustic modalities. The majority of earlier studies in the area of sarcasm identification using textual data were successful; however, a small number of studies have also been conducted with the aid of many modalities; further work is needed to enhance the results. Intuitively, sarcasm detection methods that employ deep learning techniques can automatically detect relevant information. And the deep learning approach should perform better when there are sufficient samples for training. To increase the accuracy of MSD, we propose a multi-modal fusion network using a feature fusion algorithm based on deep learning.

We experimentally evaluate our model using one of the benchmark datasets, MUStARD [11], to show that our approach is effective. As our proposed method yields superior results, hence, we infer that our proposed approach is better prepared to capture the interrelationships among various modalities for the task of MSD. The following are our most significant overall contributions:

- We introduce a deep-learning-based model for MSD that combines DistilBERT and RegNet to efficiently capture meaningful information across text, visual, and acoustic modalities in a single task,
- To check the robustness of our model so that it can effectively and reliably imitate the multi-modal depiction of textual, visual & acoustic modalities and yield ground-breaking results for MSD, we perform extensive analysis and test our proposed approach using one of the benchmark datasets, MUStARD.

This is how the rest of the manuscript will be laid out. In Section II, we will review the most recent research studies that have contributed to the advancement of sarcasm recognition, with a particular emphasis placed on multimodal sarcasm identification. The suggested novel framework is explained in

979-8-3503-3936-9/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:41 UTC from IEEE Xplore. Restrictions apply.

Section III. The experiment's results are analysed Section IV, and in Section V, we summarise our findings and offer some suggestions for future directions.

II. Related Work

Both single-modal and cross-modal sarcasm identification methods from the existing cutting-edge works are briefly discussed here.

A. Single modal

This section discusses previous research studies on sarcasm detection that have used only a single modality in their analyses.

1) Text-Based Sarcasm Detection (TBSD)

The method of predicting whether script-based data is sarcastic or non-sarcastic is known as TBSD. Machine learning-based techniques [5] have historically been used to process text for irony identification. Before the development of BERT-based models, other deep learning-based models, such as LSTM [6], attracted considerable interest because of their ability to handle script-based data. Bert-based modes [7] are widely used nowadays to identify sarcasm, emotion, and sentiment from textual data.

2) Acoustic Sarcasm Detection (ASD)

To identify sarcasm in acoustic data, researchers have mostly concentrated on finding certain auditory patterns that are linked to sarcastic utterances. Several aspects [8], [9] such as average amplitude, variations in the amplitude, the pace of the utterance, etc., are analyzed for the detection of irony in audio. Slower speech speed and higher intensity were identified by [9] as possible indicators of sarcasm. Subsequently, [10] [11], [12] examined rhythmic and spectral aspects of voice to determine sarcasm, regardless of the context.

B. Multimodal

In this section, the previous research that was done on sarcasm detection utilizing a range of different modalities is covered.

1) Sarcasm Detection using Image-Text Pairs

As of today, a great number of research have been carried out to identify how one might spot the irony in a certain social post. Users have traditionally relied on data presented in written format; but, in the present era, the use of visuals, videos, music, and emoticons is becoming more common. As a consequence of this, we often come across posts on online community forums that combine text as a caption with visual clues. Enormous amount of research studies has been in the field of sentiment recognition [13], [14] but the detection of sarcasm via the use of caption and visual content [15] [16] has been the focus of great research that is referred to as sarcasm detection using image-text pairs.

2) Sarcasm Detection Using Videos

Every day several social media platforms are bombarded with a variety of multimedia data which includes videos as well. Detecting sarcasm from these opinionated videos is very crucial. From the past few years around 2019, researchers are more eager to work in of multimodal sarcasm recognition with the aid of video datasets. [3] proposed a benchmark dataset MUStARD and analyzed it for the detection of sarcasm. [17]–[20] are a few other recent research studies that proposed cutting –edge methods for detecting sarcasm in videos.

III. Proposed Framework

In this section, the formulation of the objective, and the architecture for processing textual, visual & acoustic input are discussed in detail. The diagram in **Figure 2** illustrates the proposed framework.

A. Formulation of the objective

Given multi-modal data which contain multiple video utterances U_n, each utterance $d \in U_n$ may be written as $d = \{s, v, a, c\}$ where s represents sentences, v represents visual, a represents acoustic, and c refers to the class label. Here, n is the total number of utterances. The main goal is to classify these video utterances as sarcastic and non-sarcastic. Using d as input, MSD seeks to train a sarcasm classifier to accurately categorize multi-modal samples either as sarcasm or non-sarcasm.

B. Model to retrieve features from the textual content

We use DistilBERT [21] to retrieve features from the textual content of the utterances in the video collection. DistilBERT is a compact, quick, less expensive, and lightweight Transformer model introduced via compressing the BERT base model architecture. It has 40% fewer trainable parameters and executes 60% quicker as compared to other BERT-based models. For this reason, we have chosen [21] model to serve as our text encoder. Specifically, we took the mean of the final four layers of the initial token's transformer stack ([CLS]) to obtain a distinct text representation of the dimension $\mathbb{D}_s = 1024$.

C. Model to retrieve features from the visual content

The architecture will almost certainly remain a key aspect therefore, concentrating on broader design spaces may lead to more effective networks. Motivated by this, the RegNet [22] model has been used to retrieve relevant information from the visual content of the utterances. First, each video frame is pre-processed by scaling, normalizing, and cropping the central portion. The mean of the resultant feature vector $\mathbb{D}_v = 2048$ is then computed to provide the final representation for each video utterance.

D. Model to retrieve features from the acoustic content

Librosa, a well-known sound processing library, is one that we make use of to process the acoustic content. We begin by loading the audio sample and using a heuristic vocal-extraction approach, to eliminate extraneous signal interference. Then the processed signal is divided into equal-sized segments to obtain spatial features such as MFCC, spectrograms, spectral centroid, and their temporal derivatives. All of these retrieved characteristics are then combined to form a $\mathbb{D}_a = 1024$ dimensional combined representations.

E. Fusion strategy

To combine the textual, visual, and acoustic features obtained in the earlier steps, a feature-based fusion technique has been implemented. The final dense layer is then applied to the combined feature vector to accomplish the classification task, determining whether the multimodal utterance is sarcastic or not.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:41 UTC from IEEE Xplore. Restrictions apply.

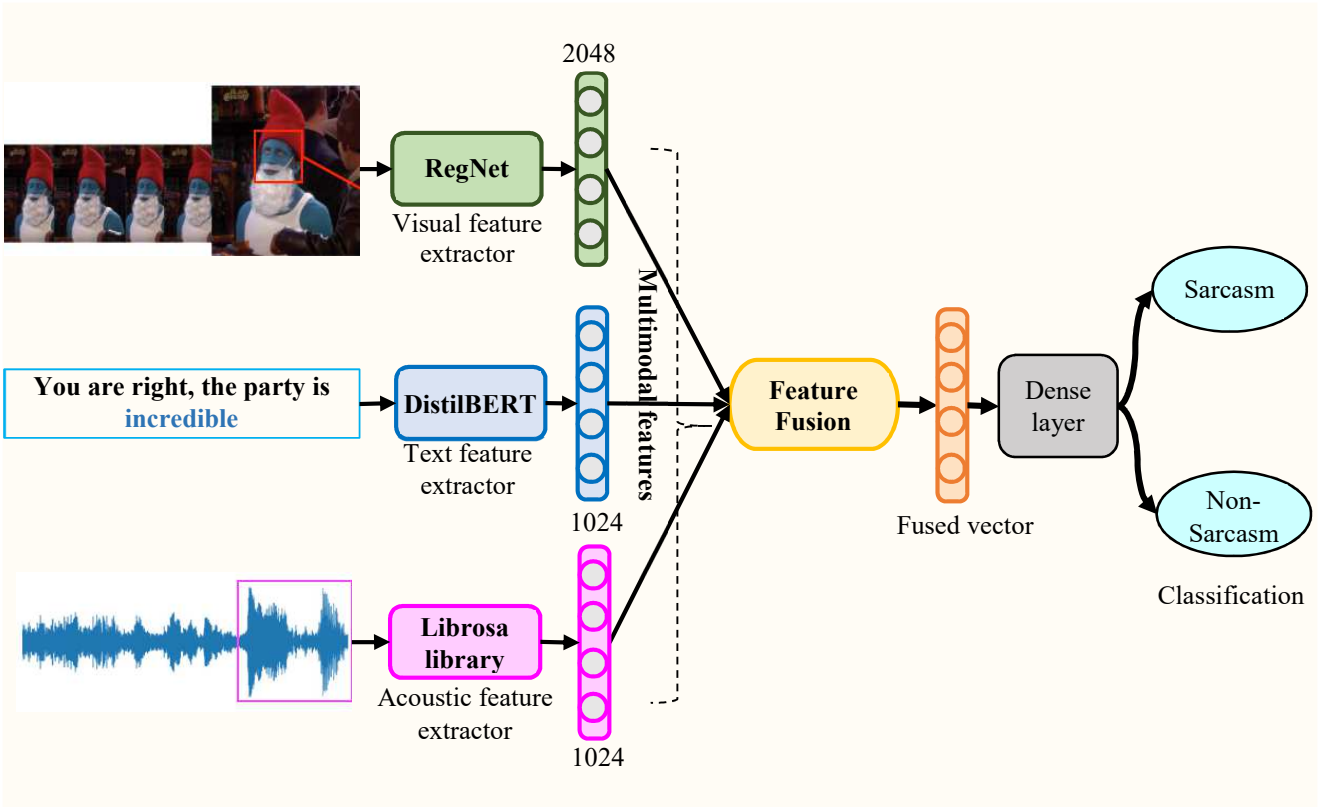


Figure 2 Proposed Framework

IV. Experimental Results

A. Dataset Used

We utilized the publicly available **MUStARD** [3] dataset for MSD, which has a total of 690 video utterances. out of 690 video utterances, 345 are sarcastic and the remaining are non-sarcastic. Our proposed model is tested in two different experimental setups i.e., speaker-independent and speaker-dependent. In the scenario of speaker-independent, all the utterance from one particular series is taken for testing purpose while others are taken for training.

B. Implementation Details

We have used DistilBERT to process textual content, RegNet to process visual data, and librosa library to process acoustic metadata from the videos. For both experimental setups, we used the same hyperparameters for training and validation. Adam optimizer and categorical cross-entropy loss function have been employed with a default learning rate of 0.001. The dimension of $\mathbb{D}_s$, $\mathbb{D}_v$, and $\mathbb{D}_a$ are successfully set to 1024, 2048, and 1024. The experimentation is run for 150 epochs. During the validation of our framework, the batch size is set to 32. To mitigate the risk of overfitting, we employ a dropout probability of 0.4.

C. Evaluation Measures

Weighted $\boldsymbol{Accuracy}$ $(\mathcal{A})$, $\boldsymbol{Precision}$ $(\mathcal{P})$, $\boldsymbol{Recall}$ $(\mathcal{R})$, $\boldsymbol{and}$ $\mathcal{F}1 - \boldsymbol{Score}$ $(\mathcal{F}1)$ are the standard metrics we employ for evaluation. Using an average of results from five separate experiments, we give our outcomes for the speaker-dependent alternative.

D. Evaluated Results

The experimental results of our MSD methodology applied to the MUStARD dataset are shown in **Table *1*** and **Table *2***. The outcomes of our research studies provide evidence that our framework is effective, and performs better than various other cross-modal approaches for MSD. It works far better than earlier methods, exceeding them in all metrics. **Figure 3** and **Figure 4** depict the year-wise comparison of the results from our proposed model to the existing cutting-edge.

Table 1 Experiment analysis on the test samples of the MUStARD dataset for speaker-dependent setting

		Speaker Dependent			
	Methods	$\mathcal{A}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}1$
Modality (text + visual + acoustic)	IWAN[2]	-	75.2	74.6	74.5
	[19]	-	73.8	73.62	73.58
	MuLOT [20]	**78.57**	-	-	-
	[3]	-	71.9	71.4	71.5
	Proposed (Ours)	78.52	**77.2**	**76.6**	**76.7**

Table 2 Experimental analysis of the test samples of the MUStARD dataset for speaker-independent setting

		Speaker Independent			
	Methods	$\mathcal{A}$	$\mathcal{P}$	$\mathcal{R}$	$\mathcal{F}1$
Modality (text + visual + acoustic)	IWAN[2]	-	71.9	71.3	70.0
	[19]	-	63.85	63.09	62.45
	MuLOT [20]	-	-	-	-
	[3]	-	64.3	62.6	62.8
	Proposed (Ours)	**76.41**	**73.3**	**72.7**	**72.8**

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:41 UTC from IEEE Xplore. Restrictions apply.

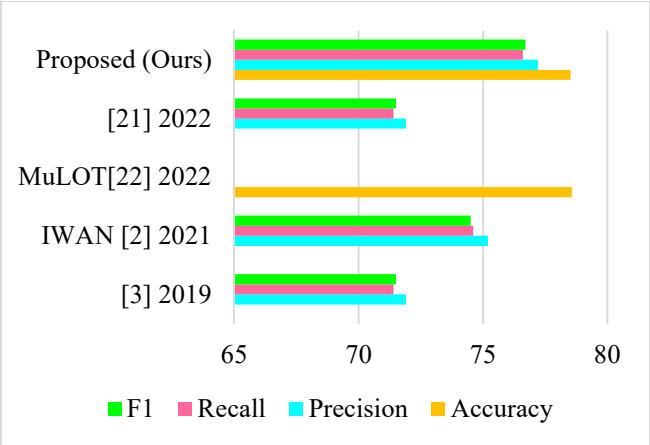


Figure 3 Year-wise methods comparison with our proposed model for speaker-dependent setting

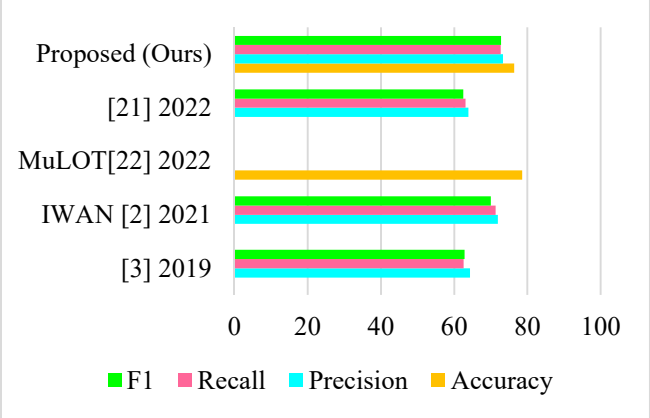


Figure 4 Year-wise methods comparison with our proposed model for speaker-independent setting

V. Conclusion and Future Scope

In this research study, a novel framework is introduced which is the combination of DistilBERT and RegNet. In general, the proposed approach includes the following three main stages: individual feature extraction from each of the separate modalities, feature fusion of all the obtained feature vectors, and classification. Our proposed model beats the baseline results which proves the robustness of our method. Apart from the feature fusion strategy we will try to cover other multimodal fusion techniques in the future such as spatiotemporal fusion, cross-attention fusion, etc. We will also try to integrate attention modules in the visual architecture to focus on the most viable information. Until now, there has been no research done in the area of visual sarcasm detection due to a lack of a dataset; as a result, we are also attempting to create such a dataset.

References

[1] R. Schifanella, P. De Juan, J. Tetreault, and L. Cao, "Detecting sarcasm in multimodal social platforms," in *MM 2016 - Proceedings of the 2016 ACM Multimedia Conference*, 2016. doi: 10.1145/2964284.2964321.

[2] Y. Wu *et al.*, "Modeling Incongruity between Modalities for Multimodal Sarcasm Detection," *IEEE Multimedia*, vol. 28, no. 2, 2021, doi: 10.1109/MMUL.2021.3069097.

[3] S. Castro, D. Hazarika, V. Pérez-Rosas, R. Zimmermann, R. Mihalcea, and S. Poria, "Towards multimodal sarcasm detection (an obviously perfect paper)," in *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 2020. doi: 10.18653/v1/p19-1455.

[4] D. S. Chauhan, D. S R, A. Ekbal, and P. Bhattacharyya, "Sentiment and Emotion help Sarcasm? A Multi-task Learning Framework for Multi-Modal Sarcasm, Sentiment and Emotion Analysis," 2020. doi: 10.18653/v1/2020.acl-main.401.

[5] A. Joshi, V. Tripathi, K. Patel, P. Bhattacharyya, and M. Carman, "Are word embedding-based features useful for sarcasm detection?," in *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2016. doi: 10.18653/v1/d16-1104.

[6] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2019.2963630.

[7] A. Baruah, K. Das, F. Barbhuiya, and K. Dey, "Context-Aware Sarcasm Detection Using BERT," 2020. doi: 10.18653/v1/2020.figlang-1.12.

[8] H. S. Cheang and M. D. Pell, "The sound of sarcasm," *Speech Commun*, vol. 50, no. 5, 2008, doi: 10.1016/j.specom.2007.11.003.

[9] P. Rockwell, "Lower, slower, louder: Vocal cues of sarcasm," *J Psycholinguist Res*, vol. 29, no. 5, 2000, doi: 10.1023/A:1005120109296.

[10] J. Tepperman, D. Traum, and S. Narayanan, "'Yeah right': Sarcasm recognition for spoken dialogue systems," in *INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP*, 2006. doi: 10.21437/interspeech.2006-507.

[11] G. A. Bryant, "Prosodic contrasts in ironic speech," *Discourse Process*, vol. 47, no. 7, 2010, doi: 10.1080/01638530903531972.

[12] J. Woodland and V. Daniel, "Context and intonation in the perception of sarcasm," *Metaphor Symb*, vol. 26, no. 3, 2011, doi: 10.1080/10926488.2011.583197.

[13] A. Pandey and D. K. Vishwakarma, "VABDC-Net: A framework for Visual-Caption Sentiment Recognition via spatio-depth visual attention and bi-directional caption processing," *Knowl Based Syst*, p. 110515, Mar. 2023, doi: 10.1016/j.knosys.2023.110515.

[14] A. Pandey and D. K. Vishwakarma, "Attention-based Model for Multi-modal sentiment recognition using Text-Image Pairs," in *2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*, IEEE, Feb. 2023, pp. 1–5. doi: 10.1109/ICITIIT57246.2023.10068626.

[15] S. Sangwan, M. S. Akhtar, P. Behera, and A. Ekbal, "I didn't mean what I wrote! Exploring Multimodality for Sarcasm Detection," in *2020 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9206905.

[16] D. S. Chauhan, D. S R, A. Ekbal, and P. Bhattacharyya, "All-in-One: A Deep Attentive Multi-task Learning Framework for Humour, Sarcasm, Offensive, Motivation, and Sentiment on Memes," 2020.

[17] S. Poria, D. Hazarika, N. Majumder, and R. Mihalcea, "Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research," *IEEE Trans Affect Comput*, 2020, doi: 10.1109/TAFFC.2020.3038167.

[18] Y. Zhang *et al.*, "CFN: A Complex-Valued Fuzzy Network for Sarcasm Detection in Conversations," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 12, 2021, doi: 10.1109/TFUZZ.2021.3072492.

[19] N. Ding, S. wei Tian, and L. Yu, "A multimodal fusion method for sarcasm detection based on late fusion," *Multimed Tools Appl*, vol. 81, no. 6, 2022, doi: 10.1007/s11042-022-12122-9.

[20] S. Pramanick, A. Roy, and V. M. Patel, "Multimodal Learning using Optimal Transport for Sarcasm and Humor Detection," in *Proceedings - 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2022*, 2022. doi: 10.1109/WACV51458.2022.00062.

[21] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.01108

[22] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020. doi: 10.1109/CVPR42600.2020.01044.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:51:41 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

Natural Ester Oil-Based High Permittivity CaCu3Ti4O12 (CCTO) Nanofluids an Eco-friendly Insulating Medium for Power Transformer

Arun Ram Prasath R T, Sankar Narayan Mahato, Nirmal Kumar Roy, Thomas Paramanandam, Mubeen Akhtar Ansari and Joges Chandra Biswas

***Abstract*- Natural ester oils are being considered as an alternative and eco-friendly insulating medium presently preferred by the power industry for power transformer application. This work deals with the development of natural ester oils like rice bran and sunflower oil-based high permittivity CCTO nanofluids in different volume concentrations. The investigation of dielectric and physico-chemical properties is performed in accordance with ASTM and IEC standards. The enhancement of AC BDV is found at 45% in rice bran oil-based CCTO nanofluid. Increment in dielectric permittivity, decrement in loss tangent, PDIV and DC resistivity are the common factors observed in both the ester oil samples. Also, flash point and thermal conductivity properties showed enhancement with an increment of volume fraction loading. Hence, due to the optimum dispersion of nano-CCTO loading at ≤0.01% in both the base oils of rice bran and sunflower oil, remarkable enhancement is achieved in both the nanofluid samples to develop an eco-friendly insulating medium for power transformer application.**

***Index Terms*-CCTO nanopowder, dielectric property, natural ester oil, physico-chemical property**

I. Introduction

NATURAL ester oil was first introduced in early 1998s for retro-filling purposes in transformers. A completely vegetable oil-based ester fluid transformer 420 kV at 400 MVA was commissioned at a German transmission network [1, 2]. In recent years, eco-friendly, clean and functional technologies are in great demand in electrical installations. Higher moisture tolerance, bio-degradability, slow aging rate, recyclability, low noise, and less initial and operating cost of ester-based insulating oils are highly preferred for transformer applications [3, 4]. Presently, Indian transformer industries have turned their attention to next-gen insulating liquids, including eco-friendly insulating oils like synthetic esters and natural esters [5]. The seed oil or vegetable-based esters are preferred due to their moderate oxidation stability compared to mineral oils, though the natural esters are lower in oxidation stability than synthetically modified esters [6]. The content of fatty acids in natural ester oil determines its performance. Sunflower oil (SFO), Rice bran oil (RBO) and Corn oil (CO) samples among others, contain a high level of fatty acids like palmitic, oleic and linoleic acids [7, 8]. In the last five decades of insulation development, anti-oxidants (natural or synthetic type) and nano-additives are majorly preferred for enhancement studies of insulating liquids [9]. Pavel et al. [10] reported that different natural and synthetic anti-oxidants mixed with different natural ester oils give better insulation results compared to mineral oils. Breakdown studies in synthetic ester and natural ester blend under AC homogeneous field exhibit higher breakdown strength compared to synthetic ester and mineral oil blend samples [11].

Recent developments in nanotechnology and its application in insulating liquids have enabled higher dielectric strength, permittivity, and thermal conductivity. The insulating and charge characteristics of TiO2 (semi-conductive type) nanoparticle dispersed in natural ester oil improves AC BDV by a factor of 1.3 [12]. Liu et al. [13] investigated high-performance Al-N nanoparticles doped with ester base oil the electrical properties like AC, lightning impulse BDV and physical properties like thermal conductivity showed enhanced performance in both electrical and physical properties than base oil. Natural esters from Honge, Neem, Mustard and Punna oils have been tested by dispersing Al2O3, BN and Fe3O4 in different volume concentrations [14]. These ester-based nanofluids not only improve breakdown voltage (BDV) property but also enhance flash and fire point properties with an increase in volume concentrations. Stable ester-based nanofluids developed from exfoliated hexagonal Boron-Nitride (h-BN) mixed with isopropyl alcohol result in 18% enhancement in AC BDV and 11.9% increment of thermal conductivity under RTP conditions [15].

Nanomaterials of Fe2O3 (conductive type) and TiO2 (semi-conductive type) based natural ester nanofluids also show enhanced dielectric breakdown strength at 0.01% volume concentration for both types [16]. The common enhancement of nanoparticles with different insulating oils was attributed to the formation of bond, bound and loose

(Corresponding author: Arun Ram Prasath R T)

Arun Ram Prasath R T, Mubeen Akhtar Ansari, and Joges Chandra Biswas are with the CSIR-National Physical Laboratory (NPL), New Delhi, India (e-mail: arunramprasathrt@nplindia.org; ansari@nplindia.org; jcbiswas@nplindia.org)

Sankar Narayan Mahato and Nirmal Kumar Roy are with the National Institute of Technology Durgapur (NITD), West Bengal, India (e-mail: sankar.mahato@ee.nitdgp.ac.in; nirmalkumar.roy@ee.nitdgp.ac.in)

Thomas Paramanandam is with the Central Power Research Institute (CPRI)-Banglore, Karnataka, India (e-mail: thomas@cpri.in)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.

© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

multi-layers encapsulated around the nanoparticle [17]. Studies of dielectric and physico-chemical properties performed in natural ester-based nanocomposite fluids showed higher dielectric and physical properties with increment in nanopowder loading. Investigations on dielectric properties and the use of classical nanoparticles are in still progress. Based on the survey, it is found that the classical nanoparticles like conductive, and semi-conductive types possess certain limitations like moderate dielectric and thermal stability nature [18, 19]. Meanwhile, high dielectric ceramic-based CaCu3Ti4O12 (CCTO) nanopowder with different flexible synthesis techniques resulted in low loss and thermally stable due to its internal barrier layer capacitance (IBLC) property, which attracted the researchers for different engineering applications [20]. Hence, high dielectric CCTO nanopowder is chosen for this research study in order to find an alternative insulation of high voltage power apparatus.

Based on this motivation, an initiative has been made, in this work, to study the electrical, physico-chemical properties of eco-friendly natural ester-based high permittivity CCTO nanofluids under different volume concentrations for suitability in power transformer applications.

II. Experimentation

A. Base Fluids

Natural ester oil is usually a seed-based oil, purchased from a leading manufacturer in India. The natural ester oils like Sunflower oil (SF) and Rice bran oil (RB) are used in this work as Base fluid-1 and Base fluid-2 and designated as SF-0 and RB-0. The fatty acid percentage and critical properties of the base fluids are given in Tables I and II respectively.

B. Preparation of CCTO Nanofluids

The preparation procedure of CCTO nanoparticles using the oxalate precursor route and TEM-based particle size analysis of the synthesized nanoparticle have been presented in our previous works [21–23]. The phase-pure CCTO nanoparticles confirm the random prevalence of semi-coherent precipitate within some of the grains resembling spherical balls-like shapes in size ranges of 30–200 nm. The elemental composition of CCTO nanopowder was confirmed by energy dispersive spectra (EDAX) analyses, as shown in Fig 1.

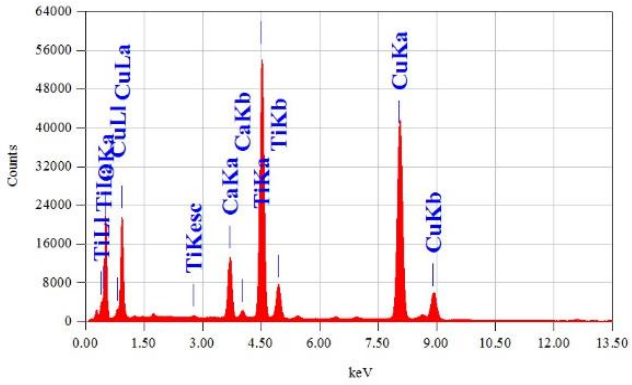


Fig. 1. Elemental composition of high dielectric CaCu3Ti4O12 (CCTO) nanopowder obtained from EDAX Spectra image

The present research work is a continuation of previous work carried out on synthetic ester and mineral oils to determine the suitability of insulating oils with high dielectric CCTO nanofluid in power transformer applications [22, 23]. In this work, refined sunflower and rice bran oils were used to prepare nanofluids with different volume percentages. Prepared nanofluid samples of sunflower oil with volume 0.001 %, 0.0025 %, 0.005 %, 0.01 % and 0.05 % were designated as SF-1, SF-2, SF-3, SF-4 and SF-5. Similarly, the prepared samples of rice bran oil with volume 0.001%, 0.0025%, 0.005%, 0.01% and 0.05% were designated as RB-1, RB-2, RB-3, RB-4 and RB-5. For the preparation of the

TABLE I

FATTY ACID PERCENTAGES OF BASE FLUIDS

Base oil samples	Fatty acid percentage		
	Saturated	Un-saturated	
		Mono	Poly
Base fluid-1 (SF-0)	11	20	67
Base fluid-2 (RB-0)	20	35	42

TABLE II

CRITICAL PROPERTIES OF BASE FLUID-1 AND BASE FLUID-2

Properties		Standards	Tested base fluid-1	Tested base fluid-2
AC BDV (kV/2.5 mm)		IEC 60156	42	40
PDIV (kV)		IEC 61294	10.6	11.2
DDF at 90°C		ASTM D 924	61 E-4	64 E-4
DC Resistivity (Ω cm) at 90°C		ASTM D 1169	1.3 E13	1.32 E13
Dielectric constant at 90°C		ASTM D 924	2.60	2.64
Viscosity (C-St)	RTP	ASTM D 445	19.20	19.22
	45°C		8.7	8.75
	90°C		5.15	5.18
IFT (milli N/m)		ASTM D 971	44	47
Flash point (°C)		ASTM D 93	230	237
Thermal Conductivity (W/mK) at RTP		ASTM D 2717	0.171	0.173
Acidity (mg KOH/g)		ASTM D 974	0.01	0.01

homogeneous nanofluid samples, an ultrasonic probe sonicator of 650 Watts with amplitude ±3 % and frequency 20-25 kHz was used to reduce agglomeration. The prepared samples were subjected to degassing at around 80 °C to eliminate the moisture in the prepared samples.

C. Dielectric Properties

The dielectric strength (BDV) of insulating oil indicates its

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

ability to resist electrical stress. For conducting the dielectric strength test, a 100 kV AC automatic BDV tester was used as per IEC 60156 standards [26, 28, 29]. Dielectric dissipation factor (DDF) and Dielectric constant were measured at 90 °C, according to ASTM D 924 standard [19, 22, 23]. The DC resistivity indicates the measure of specific resistance (resistivity) in insulating oil as per ASTM D 1169 standard [19, 22, 23]; it was performed at both Room Temperature and Pressure (RTP) and 90 °C temperature conditions. Partial discharge inception voltage (PDIV) in dielectric liquids was measured using a standard test cell, with a needle-sphere electrode combination with a gap of 50 mm being fixed to prevent breakdown as per IEC 61294 TR:1993 standard [23, 24, 27, 30].

D. Physico-Chemical Properties

IFT is a measure of interfacial tension between insulating oil and water under non-equilibrium conditions, expressed in milli N/m. It is a highly sensitive test for detecting small polar contaminants and oxidation products. Interfacial Tension (IFT) was performed using Kruss K9 Digital IFT tester, according to ASTM D 971 standards, at room temperature. A viscosity test was performed using a temperature-controlled viscosity bath with a calibrated viscometer as per ASTM D 445 standards [19, 22, 23]. The oil samples were tested with variations in temperature of 27 °C, 45 °C and 90 °C. The Flash point test is a measure of flammability of insulating oil and was performed in an automated Pensky-Martens closed cup apparatus as per ASTM D 93 standard [19, 22, 23]. A Thermal conductivity test was performed at RTP and 80 °C temperature conditions using KD2 Pro thermal conductivity tester and expressed in W/(mK) according to ASTM D 2717 standard [23]. Total acid content was measured at room temperature as per ASTM D 974 and is expressed as mg of KOH/g [19, 22, 23].

TABLE III

ELECTRICAL PROPERTIES OF NATURAL ESTER OIL BASED CCTO NANOFLUIDS

Samples	AC BDV (kV/2.5 mm)	Loss Tangent at 90 ºC	Dielectric Permittivity	DC Resistivity (ohm-cm) at 90 ºC	PDIV (kV/50 mm)
SF-0	42	61 E-4	2.60	1.3 E13	10.6
SF-1	48.5	59 E-4	2.8	1.18 E13	10
SF-2	53	58 E-4	2.77	1.18 E13	10.25
SF-3	56	55 E-4	2.83	1.19 E13	10.6
SF-4	57	54 E-4	2.78	1.22 E13	10.6
SF-5	50	54 E-4	2.75	1.2 E13	11.4
RB-0	40	64 E-4	2.64	1.32 E13	11.2
RB-1	44	65 E-4	2.78	1.2 E13	10.4
RB-2	45	63 E-4	2.84	1.21 E13	10.4
RB-3	54	60 E-4	2.85	1.21 E13	10.65
RB-4	58	59 E-4	2.90	1.22 E13	10.8
RB-5	52	57 E-4	2.8	1.22 E13	11.6

TABLE IV

PERCENTAGE INCREMENT AND DECREMENT VALUES OF ELECTRICAL PROPERTIES

	% Enhancement		% Decrement		
Samples	AC BDV (kV/2.5 mm)	Dielectric Permittivity	PDIV (kV/50 mm)	Loss Tangent at 90 ºC	DC Resistivity (ohm-cm) at 90 ºC
SF-0	0	0	0	0	0
SF-1	15.5	7.6	5.6	3.27	9.23
SF-2	26	6.5	3.3	4.91	9.23
SF-3	33.3	8.84	0	9.8	8.46
SF-4	35.7	6.9	0	11.4	6.15
SF-5	19	5.76	-7.54	11.4	7.69
RB-0	0	0	0	0	0
RB-1	10	5.3	7.14	-1.5	9.09
RB-2	12.5	7.5	7.14	1.5	8.33
RB-3	20	7.9	4.9	6.25	8.33
RB-4	45	9.84	3.57	7.81	7.57
RB-5	30	6.0	-3.5	10.9	7.57

III. RESULTS AND DISCUSSION

A. Dielectric Properties

AC BDV values obtained at RTP and the enhancement percent for both sunflower and rice bran oils are presented in Tables III and IV respectively. The percentage error plot is shown in Fig. 2. The sunflower oil-based CCTO nanofluid samples show an increasing trend of AC BDV values with respect to volume concentrations up to 0.01% (i.e., from SF-1 to SF-4) samples. Beyond that, decreasing trend in AC BDV was observed at 0.05 % (i.e., SF-5) sample. Similarly, in the case of rice bran-based CCTO nanofluid samples, the dielectric strength of the nanofluid was observed to increase with the increase of volume concentrations. It was also observed that the 0.005% and 0.01% (i.e., RB-3 and RB-4) samples show a sudden increment in BDV values.

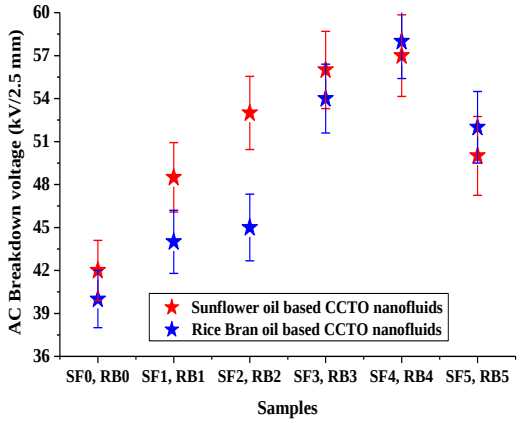


Fig. 2. Mean AC BDV values with 5% error plot of nanofluid samples from different volume concentrations

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

Rice bran oil-based CCTO nanofluid was observed to show superior enhancement of BDV value (45%) for 0.01% (i.e. RB-4) sample. Results reveal that BDV values of sunflower and rice bran oil samples at the volume concentration of 0.05% result in saturation with nanoparticle loading. It was noted that the developed nanofluid samples presented improved AC dielectric strength values resulting in charge-trapping and de-trapping processes due to the IBLC effect [25].

The DDF and dielectric permittivity values of the nanofluid samples are measured at 90 °C as per ASTM D 924 standard. The percentage decrement and increment are presented in Tables III and IV respectively. The DDF characteristics of both sunflower oil and rice bran oil-based nanofluids result in linear decreasing trends with an increase in volume concentrations as shown in Fig. 3. But compared to rice bran oil-based CCTO nanofluids, the loss tangent values of sunflower oil-based CCTO nanofluids are quite lower.

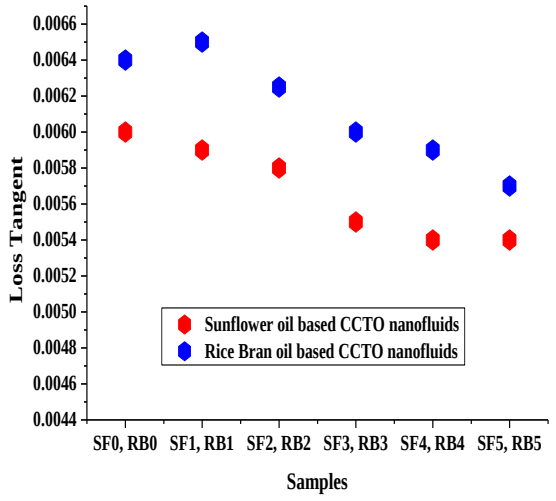


Fig. 3. DDF values of nanofluid samples from different volume concentrations

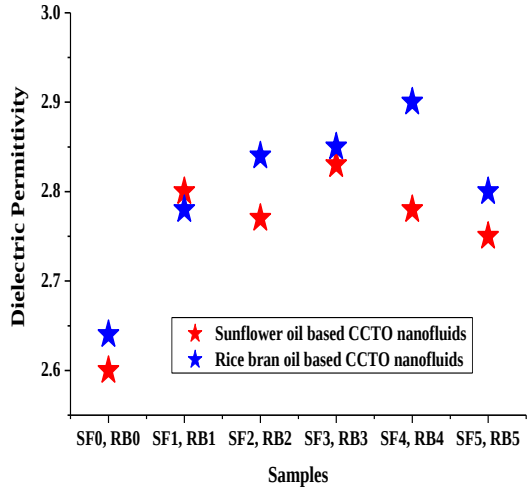


Fig. 4. Dielectric permittivity values of nanofluids samples from different volume concentrations

The sample SF-5 results in the highest decrement, which is almost the same as sample SF-4. Also, it is observed that there is a saturation of loss tangent values for sunflower oil-based CCTO nanofluid with the increase of volume concentrations as compared to rice bran oil-based CCTO nanofluid sample. This may be due to the formation of higher ionic contaminants in the case of sunflower oil-based CCTO nanofluid. From Fig. 4, it is observed that rice bran oil-based CCTO nanofluid samples show superior enhancement in dielectric constant with respect to different volume concentrations as compared to sunflower oil-based CCTO nanofluid samples. In rice bran oil samples, the steady increment of dielectric values was observed to be up to 0.01% (i.e., from RB-1 to RB-4) with the highest enhancement of 9.84% for sample RB-4. In the case of sunflower oil-based CCTO nanofluid samples, irregular improvement was observed from 0.001% to 0.005% (i.e., SF-1 to SF-3) samples with an increase in volume concentrations. Overall, it was observed that the loss tangent decreased with an increase in dielectric permittivity, probably due to the formation of nano-CCTO grain boundaries based on the optimal loading, which is similar to mineral oil-based CCTO nanofluid samples [22] and synthetic ester oil-based CCTO nanofluid samples [23].

DC resistivity values of the different samples and their percentage decrement are presented in Tables III and IV respectively. Graphical representations of DC resistivity values for both samples at RTP and 90 °C temperature conditions are presented in Fig. 5 and Fig. 6 respectively. DC resistivity values of both nanofluid samples resulted in decrement compared to base oil samples with respect to different volume concentrations. Under RTP conditions, compared to base oil samples, both nanofluid samples resulted in decrement with an increase in volume concentrations of nanoparticles. In addition, maximum decrement was observed at 0.05% (i.e. RB-5) sample as compared to other nanofluid samples. At 90 °C, the decrement in resistivity values was observed with an increase in volume concentrations. But for 0.01% and 0.05% samples, there was a slight increase in DC resistivity.

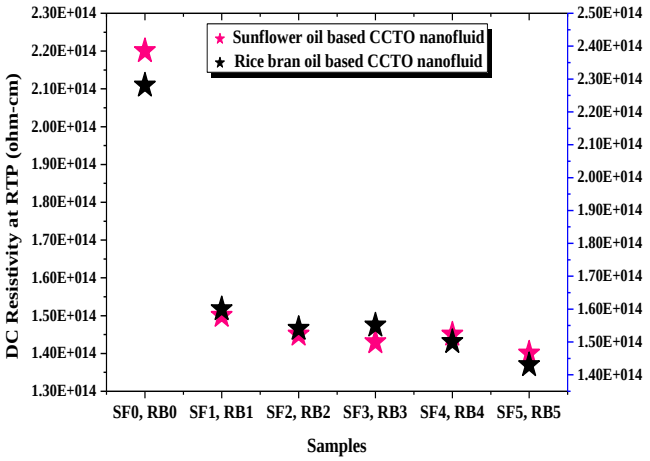


Fig. 5. DC Resistivity values of nanofluid samples from different volume concentrations at RTP conditions

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

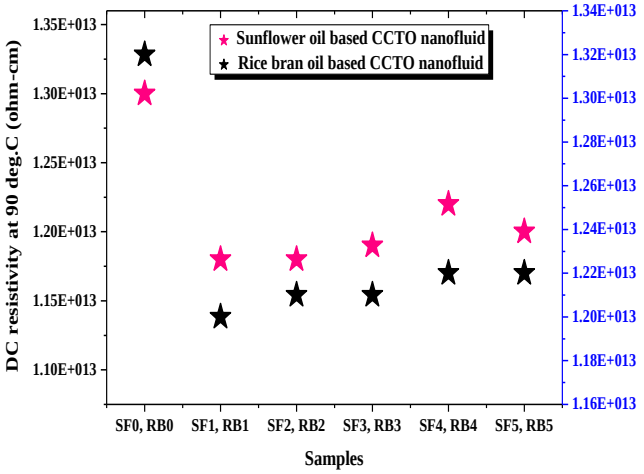


Fig. 6. DC Resistivity values of nanofluid samples from different volume concentrations at 90 ºC conditions

The highest increment was observed in the 0.01% (i.e. SF-4) sample among all the nanofluid samples. A similar trend was observed when the resistivity was measured at 90 °C with synthetic ester oil-based CCTO nanofluid samples [23].

PD inception voltage values measured under RTP conditions and the percentage decrement are given in Tables III and Table IV respectively. The percentage error plot of PDIV values corresponding to the electric field generated at needle electrode arrangement [23, 24] for base oil and nanofluid samples are presented in Fig.7 and Fig. 8 respectively.

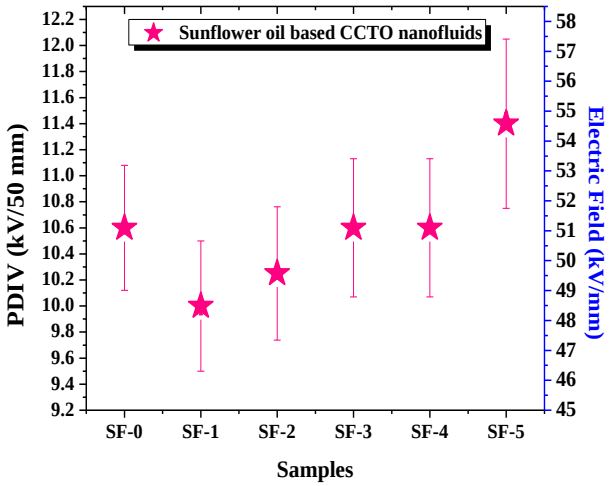


Fig. 7. PDIV values of sunflower oil-based nanofluids samples from different volume concentrations

It was observed that with an increase in volume concentrations, rice bran oil samples show reduced PDIV characteristics, similar to sunflower oil samples. But for 0.05 % (SF-5 and RB-5), the detected PDIV results are comparatively greater than the respective base oil samples. These nanomaterials resulted in a little increment of PD inception characteristics due to saturation following further addition of volume concentrations, owing to irregular size orientation of nanomaterials. Further, due to the high dielectric nature of CCTO nanoparticles, both natural ester-based nanofluid samples results minimized PDIV initiation. It was also observed that the increase of PDIV values, which was observed after certain volume concentration, is not linear with the increase of volume concentrations in both nanofluid samples. The highest decrement of PDIV value was observed for both 0.001 % and 0.0025 % (i.e. SF-1 and SF-2) samples.

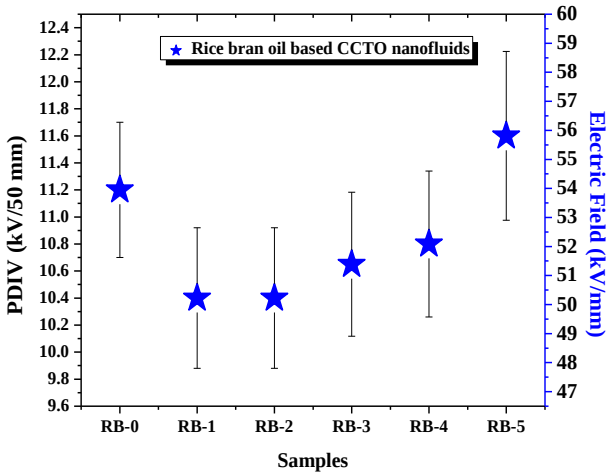


Fig. 8. PDIV values of rice bran oil-based nanofluids samples from different volume concentrations

B. Physico-Chemical Properties

Interfacial tension was measured under the RTP condition. The IFT values and their percentage decrements are shown in Tables V and VI respectively. The variation of IFT values in base oils and nanofluid samples with respect to volume concentrations is presented in Fig. 9. Both sunflower and rice bran oil-based CCTO nanofluid samples give an almost similar decreasing trend with an increase in volume concentrations.

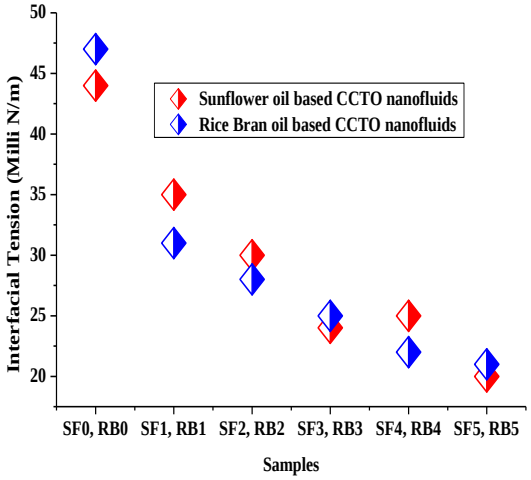


Fig. 9. IFT values of nanofluid samples from different volume concentrations

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

Analyses of comparative decrement between the nanofluid samples reveal that the highest decrement is at 0.05% volume concentration for both sunflower and rice bran oil samples (i.e. SF-5 and RB-5). The overall decrement trend was attributed to the occurrence of higher polarization at the oil-water interface region. Similar decremental nature of IFT values was observed with synthetic ester oil-based CCTO nanofluid samples [23].

TABLE V

PHYSICO-CHEMICAL PROPERTIES OF NATURAL ESTER OIL BASED CCTO NANOFLUIDS

Samples	IFT (milli N/m)	Viscosity (C-St) at 90 °C	Flash point (°C)	Thermal Conductivity (Wm/K) at 80 °C	Acidity (mg KOH/g)
SF-0	44	5.15	230	0.158	0.01
SF-1	35	5.15	252	0.162	0.01
SF-2	30	5.16	255	0.162	0.015
SF-3	24	5.16	258	0.164	0.01
SF-4	25	5.17	264	0.164	0.02
SF-5	20	5.18	265	0.165	0.02
RB-0	47	5.18	237	0.160	0.01
RB-1	31	5.18	248	0.164	0.01
RB-2	28	5.19	253	0.165	0.01
RB-3	25	5.19	258	0.165	0.01
RB-4	22	5.20	264	0.167	0.012
RB-5	21	5.21	270	0.168	0.02

TABLE VI

PERCENTAGE INCREMENT AND DECREMENT VALUES OF PHYSICO-CHEMICAL PROPERTIES

	% Enhancement				% Decrement
Samples	Viscosity (C-St) at 90 °C	Flash point (°C)	Thermal Conductivity (Wm/K) at 80 °C	Acidity (mg KOH/g)	IFT (milli N/m)
SF-0	0	0	0	0	0
SF-1	0	9.5	2.53	0	20.4
SF-2	0.19	10.8	2.53	50	31.8
SF-3	0.19	12.17	3.79	0	45.4
SF-4	0.38	14.78	3.79	100	19
SF-5	0.58	15.2	4.43	100	54.5
RB-0	0	0	0	0	0
RB-1	0	4.64	2.5	0	34.04
RB-2	0.19	6.75	3.12	0	40.4
RB-3	0.19	8.86	3.12	0	46.8
RB-4	0.38	11.39	4.37	20	53.19
RB-5	0.57	13.92	5.0	100	55.3

The Viscosity test was conducted at different temperatures. The measured viscosity values and their percentage enhancement are shown in Tables V and VI respectively. The variation of viscosity with respect to different volume concentrations at different temperatures is presented as bar chart in Fig. 10. A minor increase in viscosity was detected with an increase of volume concentrations. The viscosity was found to reduce with the increase in temperature. Among all the samples, 0.05 % volume concentration of both nanofluids (i.e. SF-5 and RB-5) samples showed superior results with enhanced viscosity to base samples. Viscosity enhancement is on account of the density of nano-CCTO powder, which is similar to that of mineral oil-based CCTO nanofluid [22] and synthetic ester oil-based CCTO nanofluid [23] samples.

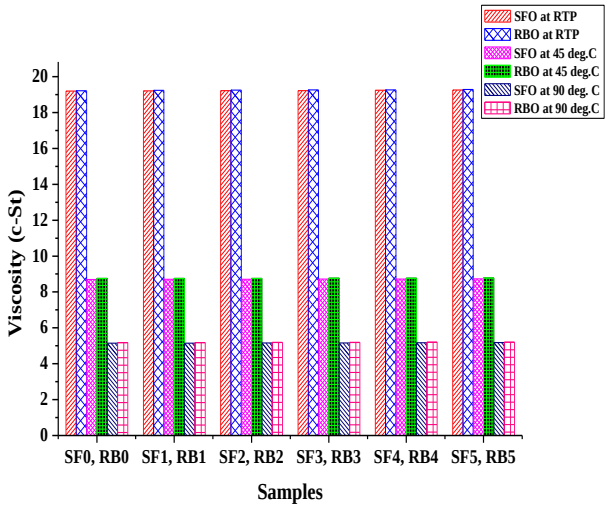


Fig. 10. Viscosity values of nanofluid samples from different volume concentrations

Flash point tests in base oils and nanofluid samples were performed as per ASTM D93 standard at RTP conditions. The flash point values and the enhancement percent are shown in Tables V and VI respectively. The variations of flash points of natural ester oil samples are graphically represented in Fig. 11.

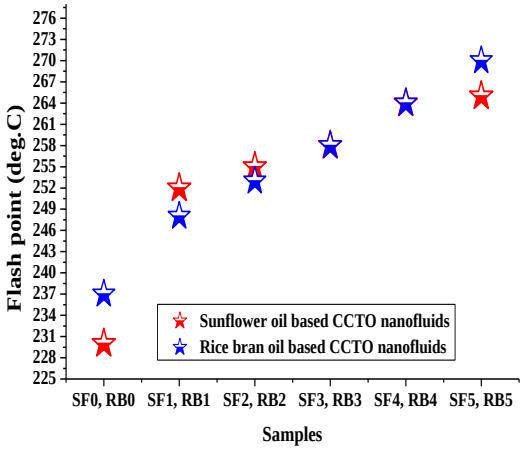


Fig. 11. Flash point values of nanofluid samples from different volume concentrations

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

It was observed that the flash point increases almost linearly with the increase of volume concentrations for both samples. It was also observed that both nanofluid samples resulted in enhanced flammability characteristics to base oil samples. Moreover, the natural ester oil-based CCTO nanofluid resulted in improved flammability characteristics, similar to mineral oil-based CCTO nanofluid samples [22] and synthetic ester oil-based CCTO nanofluid samples [23].

Thermal conductivity tests in base oils and nanofluid samples were performed at RTP and at 80 °C respectively. The test results and the percentage increment are presented in Tables V and VI respectively. At both temperatures, the thermal conductivity increased with an increase in volume concentrations of nano-CCTO powder, which is presented in Fig. 12.

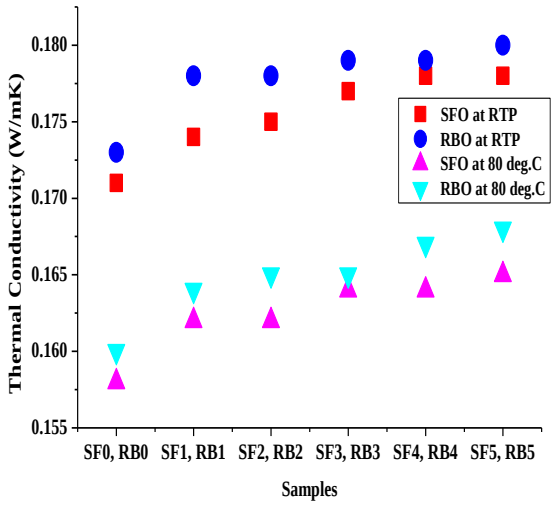


Fig. 12. Thermal conductivity values of nanofluid samples from different volume concentrations

Moreover, the thermal conductivity reduced with the increase in temperature. It is found that the percentage increments of thermal conductivity for samples SF-5 and RB-5 are 4.43% and 5% at 80 °C respectively. The thermal conductivity enhancement in natural ester oil-based CCTO nanofluid depends on the presence of nanoparticle size, shape, and grain boundaries. A similar trend was noted in synthetic ester oil-based CCTO nanofluid [23].

The acidity test was carried out at RTP as per ASTM D 974 standard. The observed values are tabulated and the enhancement percentages are given in Tables V and VI respectively. The variation of acidity with the increase of different volume concentrations is presented in Fig. 13. Acidity was seen to be very low for natural ester oil-based nanofluids, as well as for a few nanofluid samples. It was also observed that the acidity increased with the increase of volume concentrations in an acceptable range (i.e., ≤0.05 mg KOH/g). Further, nanofluid samples may result in the generation of polar compounds under aging conditions [22, 23].

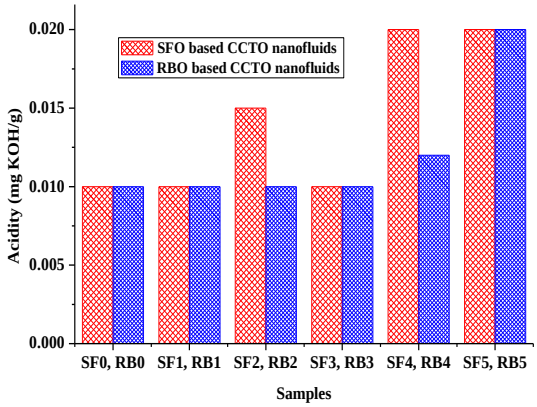


Fig. 13. Acidity values of nanofluid samples from different volume concentrations

IV. Conclusion

Natural ester oil-based CCTO nanofluids have been successfully developed and critical properties like dielectric and physico-chemical properties were examined in accordance with ASTM and IEC standards. The conclusions are as follows:

• In dielectric properties, AC BDV is enhanced as the volume concentration of CCTO increases in both sunflower and rice bran oil samples, indicating that the use of high permittivity ceramics may tend to hinder breakdown occurrence by converting faster electrons to slower ones due to IBLC effect. An increment in dielectric permittivity in rice bran oil-based nanofluid samples with a reduction of loss characteristics in sunflower oil-based nanofluid samples was observed. This is attributed to the inclusion of optimized addition of nano-CCTO particles. In DC resistivity and PDIV characteristics, a decrement was observed with respect to the increase in volume concentrations.

• Viscosity in nanofluid samples results in negligible increment with variation in temperature compared to base oil samples. Interfacial tension (IFT) test results show a tremendous decrement with an increase in volume concentrations due to the high rise of polar contaminants. Thermal conductivity as well as flash and fire points, result in enhanced values in both sunflower and rice bran oil-based nanofluid samples with an increase in volume concentrations. Moreover, the acid number in nanofluid samples is in the lower range. Nanofluid samples may result in the generation of polar compounds under aging conditions.

Therefore, the developed natural ester oils-based CCTO nanofluids can enhance critical properties compared to base natural ester oil samples. The dielectric and physico-chemical properties of high dielectric CCTO-doped sunflower oil and rice bran oil-based nanofluid result in almost similar nature as compared to synthetic ester oil [23], and higher than mineral oil [22] based high dielectric CCTO nanofluid samples for power transformer application. Further, the particle agglomeration along with moisture content studies are yet to perform as a future study.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the management of CSIR-National Physical Laboratory (NPL), National Institute of Technology Durgapur (NITD) & Central Power Research Institute (CPRI-Banglore). Authors from CSIR-NPL would like to thank the Director, CSIR-NPL for his constant support.

REFERENCES

[1] Siemens Transformer-a case study: 420 kV transformer with natural ester insulation. Available https://www.siemens.com

[2] IEEE Guide for Acceptance and Maintenance of natural ester fluids in Transformer, IEEE Guide C57.147 (Revision of IEEE Std C57.147-2008), 2018-02-15.

[3] R Pillai, Firoz Havaldar and Chintamani Chitnis, "Natural esters for life and capacity enhancement of distribution transformers," *CIRED, Open Access Proc. J*, vol. 2017, pp. 355-358, 2017.

[4] T. V. Oommen, C. C. Claiborne, E. J. Walsh, and J. P. Baker, "A new vegetable oil-based transformer fluid: development and verification," *Annu. Rep. Conf. Electr. Insul. Dielect. Phenom. (CEIDP)*, vol. 1, pp. 308-312, 2000.

[5] B. Pahlavanpour, G. Wilson and R. Heywood, "Environmentally friendly insulating oil," *IEE Seminar on Engineering and the Environment - How It Affects You*, pp. 5/1-5/4, 1999.

[6] A. Ortiz, F. Delgado, Félix Ortiz, Inmaculada Fernández and Agustín Santisteban, "The aging impact on the cooling capacity of a natural ester used in power transformers," *Appl. Therm. Eng.*, vol. 144, pp. 797-803, 2018.

[7] Dhorali Gnanasekaran and Venkata Prasad Chavidi, "Vegetable oil: an eco-friendly liquid insulator", *Vegetable oil-based bio-lubricants and transformer fluids: applications in power plants*, Materials Forming, Machining and Tribology, Springer, 2018, pp. 101-121.

[8] Shengwei Cai, Cheng Chen, Huihao Guo, Shaobing Chen, Zandong Zhou, and Zhaofeng Guo, "Fire resistance test of transformers filled with natural ester insulating liquid," *14th IET International Conference on AC and DC Power Transmission (ACDC)*, pp. 1-5, 2018.

[9] J.E. Contrerasa, E.A. Rodriguez, and J. Taha-Tijerinaca, "Nanotechnology applications for electrical transformers - a review," *Electr. Pow. Syst. Res.*, vol. 143, pp. 573-584, 2017.

[10] Pavel Totzauer, Pavel Trnka, Jaroslav Hornak, Petr Kadlec and Josef Pihera, "Antioxidant variations in the nature ester oil," *18th International Scientific Conference on Electric Power Engineering (EPE)*, pp. 1-6, 2017.

[11] Grzegorz Dombek and Jaroslaw Gielniak, "Fire safety and electrical properties of mixtures of synthetic ester/mineral oil and synthetic ester/natural ester," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 5, pp. 1846–1852, October 2018.

[12] Munawar Khalila, Badrul Mohamed Janb, Chong Wen Tongc and Mohammed Ali Berawi, "Advanced nanomaterials in oil and gas industry: Design, application and challenges," *Appl. Energy*, vol. 191, pp. 287-310, 2017.

[13] Donglin Liu, Yuanxiang Zhou, Ying Yang, Ling Zhang and Fubao Jin "Characterization of high-performance Al-N nanoparticle-based transformer oil nanofluids," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 5, pp. 2757-2767, October 2016.

[14] R. Madhavan, S. Senthil Kumar and M. Willjuice Iruthyarajan, "A comparative investigation on effects of nanoparticles on characteristics of natural esters-based nanofluids," *Colloids Surf. A*, vol. 556, pp. 30–36, 2018.

[15] Wei Yao, Zhengyong Huang, Jian Li, Liya Wu, and Chenmeng Xiang, "Enhanced electrical insulation and heat transfer performance of vegetable oil-based nanofluids," *J. Nanomater*., Article ID 4504208, pp. 1-12, 2018.

[16] Victor Antonio Primo, Belen Garcia, and Ricardo Albarracin, "Improvement of transformer liquid insulation using nanodielectric fluids: A review," *IEEE Electr. Insul. Mag.*, vol. 34, no. 3, pp. 13-26, May/June 2018.

[17] Furkhan Ahmad, Asfar Ali Khan, Qasim Khan, and Md Rashid Hussain, "State-of-art in nano-based dielectric oil: a review," *IEEE Access*, vol. 7, no. 6, pp. 13396-13410, 2019.

[18] D. Vigneshwaran, M. Ravindran and S. Senthil Kumar, "Investigation of critical properties of the highly mono-unsaturated natural esters with nanocomposites," *IEEE Intl. Conf. on Innovations in Information Embedded and Communication Systems (ICIIECS)*, pp. 1–6, 2015.

[19] R.T. Arun Ram Prasath, M. Willjuice Iruthayarajan, and R. Karthik, "Performance studies on dielectric and physical properties of eco-friendly based natural ester oils using semi-conductive nanocomposites for power transformer application," *IET Sci. Meas. Technol.*, vol. 12, no. 3, pp. 323–327, 2018.

[20] I. Norezan, A.K. Yahya and M.K. Talari, "Effect of (Ba0.6Sr0.4) TiO3 (BST) doping on dielectric properties of CaCu3Ti4O12 (CCTO)," *J. Mater. Sci. Technol*., vol. 28, no. 12, pp. 1137-1144, 2012.

[21] P. Thomas, K. Dwarakanath, K.B.R. Varma and T.R.N. Kutty, "Nanoparticles of the giant dielectric material CaCu3Ti4O12 from a precursor route," *J. Phys. Chem. Solids*, vol. 69, pp. 2594–2604, 2008.

[22] R.T. Arun Ram Prasath, N.K. Roy, S.N. Mahato and P. Thomas, "Mineral oil based high permittivity CaCu3Ti4O12 (CCTO) nanofluids for power transformer application," *IEEE Trans. Dielectr. Electr. Insul*., vol. 24, no. 4, pp. 2344–2353, August 2017.

[23] P. Thomas, Nandini E. Hudedmani, R.T. Arun Ram Prasath, Nirmal Kumar Roy and Sankar Narayan Mahato, "Synthetic ester oil based high permittivity CaCu3Ti4O12 (CCTO) nanofluids an alternative insulating medium for power transformer," *IEEE Trans. Dielectr. Electr. Insul.*,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Dielectrics and Electrical Insulation. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDEI.2023.3301465

vol. 26, no. 1, pp. 314-321, February 2019.

[24] L. Calcara, M. Pompili and F. Muzi, "Standard evolution of partial discharge detection in dielectric liquids," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 1, pp. 2–6, February 2017.

[25] Mohsen Ahmadipour, Mohd Fadzil Ain and Zainal Arifin Ahmad, "A short review on copper calcium titanate (CCTO) electro ceramic: synthesis, dielectric properties, film deposition, and sensing application," *Nano-Micro Letters*, vol. 8, no. 4, pp. 291-311, October 2016.

[26] Arun Ram Prasath R. T., Mubeen Akhtar Ansari., Thomas Paramanandam., Sankar Narayan Mahato and Nirmal Kumar Roy, "Performance studies on mineral oil and natural ester oil based high dielectric CCTO nanofluids for high voltage application", *5th IEEE Conf. on CATCON*, pp123-126, Dec. 2021.

[27] IEC 61294TR: 'Determination of the Partial Discharge Inception Voltage (PDIV)-Test Procedure', 1993.

[28] Thomas. P, "Breakdown Voltage and Gassing Tendency of Synthetic Esters Based MgO Nanofluids", *4th IEEE Conf. on CATCON*, pp. 1–3, 2019.

[29] S. A. Khan, M. Tariq, A. A. Khan, and B. Alamri, "Effect of Iron/Titania-Based Nanomaterials on the Dielectric Properties of Mineral Oil, Natural and Synthetic Esters as Transformers Insulating Fluid," *IEEE Access*, vol. 9, pp.168971-168980, 2021.

[30] S. Chandrasekar, J. Chandramohan, G. C. Montanari and P. Uthirakumar, "Developing eco-friendly nanostructured oil: Partial discharge and breakdown voltage characterization of transformer corn oil," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 27, no. 5, pp. 1611-1618, Oct. 2020.

Arun Ram Prasath R T is a Scientist in the AC High Voltage and High Current Metrology, Electrical & Electronics Metrology Division, CSIR-National Physical Laboratory, New Delhi, India. He obtained his Bachelor's (Electrical) and Master's (High Voltage Engineering) degrees from Anna University, Tamilnadu, and Ph.D. degree from the National Institute of Technology Durgapur, West Bengal. His areas of interest are AC high voltage and high current metrology parameters, power systems, high voltage, insulation engineering, and nano dielectrics.

Sankar Narayan Mahato is a Professor and Head of the Department of Electrical Engineering, National Institute of Technology Durgapur, West Bengal, India. He obtained his B.E. (Electrical) degree from the National Institute of Technology Durgapur, M. Tech. (Electrical) degree from the National of Technology Durgapur, and Ph.D. degree from IIT Roorkee. His area of interest includes machine drives, power systems, and high-voltage, insulation engineering.

Nirmal Kumar Roy is a Professor (HAG) in the Department of Electrical Engineering, National Institute of Technology Durgapur, West Bengal, India. He obtained his B.E. (Electrical) degree from the National Institute of Technology Durgapur, M.Tech (Electrical) degree from the Indian Institute of Science (IISc), and Ph.D. degree from the University of South Australia. He is a recipient of the Technology Angel Award from the Department of Science and Technology. His area of interest includes power systems, high voltage, and insulation engineering.

Thomas Paramanandam is an Additional Director and Head of the Department in Dielectric Materials Division at Central Power Research Institute, Bangalore, India. He obtained his B.Sc. (Chemistry) from Madras University, M.Sc. (Analytical Chemistry) from the University of Madras, and Ph.D. (Materials Science) from the Indian Institute of Science (IISc). He has published more than 80 papers in Conferences and Journals. His areas of interest include nanofluids, high-K dielectrics, polymer nanocomposites, fuel cells, and insulation engineering.

Mubeen Akhtar Ansari is a Senior Principal Scientist and Head of the AC High Voltage and High Current Metrology, Electrical & Electronics Metrology Division, and Quality manager, CSIR-National Physical Laboratory, New Delhi, India. He obtained his B.E. (Electrical) from Nagpur University. He has contributed to the design and development of superconducting magnets for various applications, besides R&D on High-temperature superconductors. He also contributed to LF impedance and DC metrology at CSIR-NPL. His areas of interest are AC high voltage and high current metrology parameters, power systems, and insulation engineering.

Joges Chandra Biswas is a Chief Scientist and Head of the Electrical & Electronics Metrology Division, CSIR-National Physical Laboratory, New Delhi, India. He obtained his B. Tech (Hons) in Electrical from IIT Kharagpur and M.E in Electrical Engineering from Delhi College of Engineering, New Delhi. His areas of interest are AC power and energy, high voltage and high current metrology parameters, power systems, and insulation engineering.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Proceedings of the 5th International Conference on Inventive Research in Computing Applications (ICIRCA 2023)
IEEE Xplore Part Number: CFP23N67-ART; ISBN: 979-8-3503-2142-5

Natural Language Processing based Automatic Making of Use Case Diagram

Prof. Shailender Kumar
Computer Science and Engineering Department
Delhi Technological University
Delhi, India
shailenderkumar.verma@gmail.com

Aryaman
Computer Science and Engineering Dept
Delhi Technological University
Delhi, India
aryamanlakers@gmail.com

Aryan
Computer Science and Engineering Dept
Delhi Technological University
Delhi, India
aryans07024@gmail.com

Divyank Yadav
Computer Science and Engineering Dept
Delhi Technological University
Delhi, India
Yadavdivyank3@gmail.com

2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) | 979-8-3503-2142-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIRCA57980.2023.10220849

***Abstract*—In the field of software engineering, the creation of use case diagrams is crucial for effectively capturing system requirements. The current traditional development approach utilizes textual functional requirements, however, it is time-consuming and error-prone. In software engineering, our research suggests a novel way for automating the production of use case diagrams. It is well-recognized that conventional methods using written functional requirements are time-consuming and prone to inaccuracy. Our approach takes advantage of natural language processing methods for analyzing and obtaining data from natural language specifications in order to address this. This study creates use case diagrams by using predetermined rules and heuristics, which speed up the process and lowers error rates. Implementation as well as evaluation against actual requirements show that the proposed automated approach is superior to existing methods. Our approach increases efficiency and accuracy while requiring less time and effort to create diagrams while maintaining consistency. By presenting a more effective and dependable method of collecting system requirements, this research makes a contribution to the discipline of software engineering.**

***Index Terms*—Natural language processing, UML model, Use case diagram, SRS document, Software requirements**

I. Introduction

Software requirements are frequently written in standard English, which is frequently unclear, ambiguous, and unstructured. The author's mood and the environment, as well as society and area demographics, might influence the natural language used. Automating the entire process and computerizing it can be a highly effective way to reduce errors brought on by human error. Natural language documents can contain even the smallest mistake or error, which can take time to correct and have an effect on important sectors like the Internet.

A key factor in the accomplishment of software development projects is the process of determining software requirements. But when these criteria are written in normal English, difficulties like ambiguity and a lack of organization are frequently introduced. In this essay, our research examines the benefits and problems of software requirements, as well as the potential remedies provided by automation and a better comprehension of the underlying problems.

Understanding the issue and how the software will affect the product should be the first step in the software requirement process. It can be challenging to identify the demands and specifications of the software since customers frequently have trouble understanding the issue description. Writing software specifications in normal English might lead to ambiguity and confusion. Numerous elements, such as the author's mood, surroundings, society, and local demography, have an impact on the language chosen. These elements make it even harder to provide exact information and can result in mistakes in the finished output. In texts written in natural language, even the smallest error can have a big impact. The written descriptions and diagrams that form the foundation of the procedure for developing software must be carefully examined to guarantee that the system operates as planned [1]. The method of automating software requirements has many advantages. In order to examine and analyze requirements, researchers can now use automated or semi-automated procedures, saving important time. Due to the inherent ambiguity of language, previous solutions such as grammar-based approaches that depended on extracted text have been shown to be erroneous. As a result, there is an increasing trend toward automating requirements analysis using sophisticated natural language processing abilities. Software development teams can increase the effectiveness and efficiency of the requirement collection and analysis phase through automation.

II. Overview

Natural language processing (NLP) plays a crucial role in the extraction of pertinent information from texts written

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:42:22 UTC from IEEE Xplore. Restrictions apply.

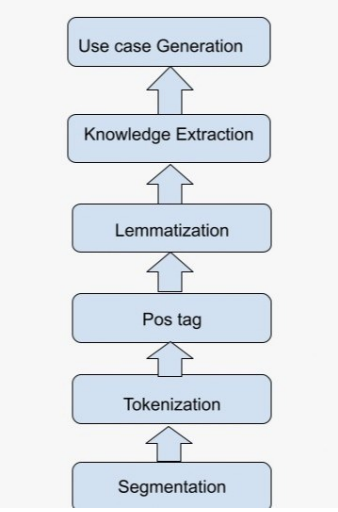


Fig. 1. Step of the Approach

in natural language, and this information can then be used directly by a machine for interpretation. As each case associated with natural language texts can be ambiguous, a wide range of NLP techniques are used to overcome this challenge. Traditionally, techniques such as use-case-driven assistant tools and language analysis for design purposes were used, but they have become outdated due to their limitations in bridging the gap between the processes involved in system design. Hence, NLP has become an essential tool in reducing the gap between the design processes.

One of the significant uses of UML use case diagrams in software development is to illustrate how a system interacts with external factors such as a company, device, or customer. The use of diagrams is also helpful in defining the system's capabilities and its interactions with different stakeholders. A use case diagram is produced to represent the system's behaviors with respect to external actors when the requirements for the system are first determined. The text can be divided into parts of speech, and each part of speech has its own unique structure. The above-mentioned technique is more accurate and efficient as grammatical and spelling errors can be corrected at an early stage with minimal effort. This approach is commonly used in natural language processing to create use cases and alternative use case diagrams.

III. Related Work

In recent years, a range of natural language processing techniques has been employed for software development. One of the earliest studies was conducted at the DEl Institute of Technology, where sentences were divided into newer sentences based on conjugation to identify different types of standard structures.

Automated tools such as R-Tool, NL-OOPS, and CM-BUILDER are utilized in natural language processing to visualize rough diagrams from user documents, which need to be further rectified by an observer [10]. Researchers have looked into using language to categorize different texts, including GOOAL, CM-Builder, and LIDA. The creation of class diagrams and the identification of use cases by comprehending their behavior through language provide a novel approach to classifying objects and building use case models. They do not, however, come without restrictions and problems that need to be taken into account. The potential for errors in diagrams created from user documents is a noteworthy limitation. Although these technologies can depict basic schematics, the natural complexity and diversity of spoken language necessitate human observation and correction. The management of many customer documents and languages presents a substantial additional barrier. Another aspect to be considered is the performance and scalability of automated programs. The capability of these applications to manage and analyze enormous datasets quickly becomes crucial as the amount of user papers rises.

Natural language processing methods include the Stanford Parser, WordNet, Stanford Named Entity Recognizer, and Java RAP, among others. The Stanford Named Entity Recognizer pinpoints the placement and structure of a word, the Stanford Parser determines sentence organization, WordNet maintains linguistic information for many sorts of identifiers, and Java RAP is useful for analyzing complicated documents and producing use-case models. These tools are only a few of the numerous natural language processing methods that are readily available.

IV. Observations

Before any software is created, the client provides the software development company with the initial requirements. The initial request from the client is delivered via a software requirements specification (SRS) document [2]. Whereas the non-functional requirements specify the system's quality, the functional requirements specify the function that the system must do.

The functional requirements, which are usually expressed in English as noun and verb phrases are commonly written in natural language. These natural language requirements are turned into a use case diagram to reduce confusion throughout the software development phase. We must discover the three components of actors, user cases, and the interaction between them in order to create a use case diagram. They can be found by making the subsequent findings.

1. Figuring out the Actors: The term actor represents any user who is playing some role in the system. They can be identified by examining who the system is made for, who is using the system, and who is making changes to the system. These are mostly the subject of the NP (Noun Phrase). So to find the actors of the system model need to find the nouns and pronouns in the SRS. Actors can be identified by examining the system's intended users, users of the system, and those who make changes to the system.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:42:22 UTC from IEEE Xplore. Restrictions apply.

2. Figuring out the Use Case: A use case explains the way an actor communicates with the framework and the sequence of events that take place there. Therefore, use cases are linked to the actors mentioned in the requirements. Therefore, the user case is most likely the VP (Verb Phrase) in the sentence so to identify the use case model need to find the corresponding VP in the NP+VP combination.

3. Figuring out the linkage between the Actor and Use case: As discussed above after finding out the actor and use case model need to find which actor is related to which use case. We are enough to find if there is any preposition connecting the two.

Our method makes the assumption that the stated needs are practical, written in English, and correctly spelled and grammatically. Then, using a six-step approach, segment, tokenize, tag POS data, lemmatize it, and extract knowledge. By identifying actors, use cases, and their interactions, a use case diagram is ultimately produced.

V. Proposed Approach

The proposed approach comprises several steps, beginning with segmentation and culminating in the creation of a use case diagram. However, it may be beneficial to include a spelling check feature to correct any spelling errors prior to commencing the process. The first step, segmentation, involves separating paragraphs into individual sentences. Next, tokenization is employed to break down each sentence into individual words, which are then labeled with Parts of Speech (POS) tags [18]. After POS tagging, then divide the POS tags into chunks. Finally, extract details from the base words to construct the use case diagram.

This text is preprocessed to prepare it for segmentation. The entire text is combined into one paragraph, and line breaks are replaced with white space. Any prefixes such as Mr. or Ms., website addresses such as .in or .com, decimal numbers, and abbreviations such as U.S. or pvt. are replaced with a special character sequence to prevent confusion with sentence-ending periods.

A. Segmentation

By breaking down the text into more manageable chunks, the segmentation process helps to solve the difficulty of comprehending the SRS. It does this by dividing paragraphs into sentences. [21]. The quantity of segments is determined by how many sentences make up a paragraph. For example, users can buy items with cash. Moreover, he can use the card. After the segmentation, the text is changed to two sentences, the user can buy with cash and Moreover, and he can buy it by card. To do the segmentation, find full stops (“.”), question marks (“?”), explanation marks (“!”) in the preprocessed text and replaces them with the ¡stop¿ to divide them into sentences. In the next process, each item would be divided into words.

B. Tokenization

Sentences are divided into tokens during the tokenization process after being produced through segmentation. It

Segmentation

User A and User B can equally connect to WiFi direct and search available users.
User A can send invitation to connect and user B can grant invitation.
Also, they can send multimedia and receive multimedia such as message, audio, video and file.
User A is same as User 8 can terminate connection.

Fig. 2. Output of Segmentation

is straightforward to use POS tagging, which assigns the keywords to a category when the sentence is converted into a collection of tokens [21]. Each word has commas and full stops, question marks, exclamation signs, and white space divide words. So split up words based on them. For example, when tokenization is applied to the given sentence, the user can buy items with cash, the output would be a group of words: user, can, buy, items, with, cash. All the symbols mentioned above can be first converted to spaces and then the text can be segregated using the location of the spaces. So Python has an inbuilt function called split() which splits the sentences based on delimiters to words, shown in Fig 3 and Fig 4.

```
def tokenization(doc):

    doc = segmentation(doc)
    tokenized_text = []
    for line in doc:
        tokenized_text.append(line.split())
    return tokenized_text
```

Fig. 3. Code snippet of tokenization

Tokenisation

User/ A/ and/ User/ B/ can/ equally/ connect/ to/ WiFi/ direct/ and/ search/ available/ users./
User/ A/ can/ send/ invitation/ to/ connect/ and/ user/ B/ can/ grant/ invitation./
Also,/ they/ can/ send/ multimedia/ and/ receive/ multimedia/ such/ as/ message,/ audio,/ video/ and/ file./
User/ A/ is/ same/ as/ User/ 8/ can/ terminate/ connection./

Fig. 4. Tokenization output

C. POS Tagging

The speech is made up of several speech components, including verbs, nouns, adjectives, and so forth. Each created token is given a POS (Part of Speech) by the POS tagger. Identifying between words with the same spelling but distinct meanings is the main goal of POS tagging. For instance, the meaning of the word right has a different context

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:42:22 UTC from IEEE Xplore. Restrictions apply.

when we say ”take a right turn” and ”you are right.” So when I use an NLTK package to generate tags of words: text=word_tokenize (“you are right, we should take a right turn”), the tags for both the right are different as their context in the sentence is different.

A word in the corpus is given a speech component based on its meaning and context during the POS tagging process [20]. Rule-based taggers and stochastic taggers are the two categories into which POS taggers fall. Whereas stochastic taggers use frequency or probability as a parameter for word tagging, Brill’s tagger uses a rule-based approach that takes context into account. The most common tag in the training corpus collection might be used to tag the term as a workaround, however, this could lead to an unfavorable order of tags.

The n-gram model, which assesses the likelihood that a tag would emerge with the n preceding tags, is an alternative to POS tagging. A stochastic method using transition and emission probabilities is an HMM with POS tagging [19]. The possibility that one tag will follow another is known as the transition probability, and the set of probabilities for each tag emitting a particular term is known as the emission probability. The proposed method relies on the SpaCy library, which was created primarily for the use of natural language processing (NLP) training. SpaCy offers numerous model ranges, each with its own special characteristics. For the English language, use the en core web sm model. To load all NLP models, use the command spaCy. load(en core web sm), which creates an NLP object and assigns parts-of-speech (POS) tags [10].

Each word in a document should be given a part-of-speech (POS) tag, which is a crucial step in natural language processing, particularly in software engineering. Token objects in the SpaCy library give users access to POS tags, which offer detailed data on word usage. Grammatical rules can reveal more information about word usage, such as whether a word serves as both the subject and a noun. A crucial step in natural language processing, particularly in software engineering, is POS labeling, which helps to distinguish between words with matching spellings but alternate meanings.

```
def posTag(doc):

    doc = segmentation(doc)
    nlp = spacy.load('en_core_web_sm',disable=['ner','textcat'])
    tagged_text = []
    for line in doc:
        sub = []
        t = nlp(line)
        for token in t:
            if token.pos_ == 'PUNCT':
                continue
            sub.append(token.text+'->'+token.pos_)
        tagged_text.append(sub)
    return tagged_text
```

Fig. 5. Code Snippet of Pos Tagging

Pos Tagging

User->PROPN; A->PROPN; and->CCONJ; User->PROPN;
B->PROPN; can->AUX; equally->ADV; connect->VERB; to->ADP;
WiFi->PROPN; direct->ADJ; and->CCONJ; search->VERB;
available->ADJ; users->NOUN;
User->PROPN; A->PROPN; can->AUX; send->VERB;
invitation->NOUN; to->PART; connect->VERB; and->CCONJ;
user->NOUN; B->PROPN; can->AUX; grant->VERB;
invitation->NOUN;
Also->ADV; they->PRON; can->AUX; send->VERB;
multimedia->NOUN; and->CCONJ; receive->VERB;
multimedia->NOUN; such->ADJ; as->ADP; message->NOUN;
audio->ADJ; video->NOUN; and->CCONJ; file->NOUN;
User->PROPN; A->PROPN; is->AUX; same->ADJ; as->SCONJ;
User->PROPN; 8->NUM; can->AUX; terminate->VERB;
connection->NOUN;

Fig. 6. Output of Pos Tagging

D. Lemmatization

In natural language processing algorithms, it is crucial to ensure that all tokenized words are transformed to their base or root form. In the English language, a single word can have multiple variations, but for our purposes, consider all forms as a single entity. For example,” run” can take on various forms such as ”running,” ”runs,” and ”ran,” but the aim is to represent all of these forms as ”run.” This process is known as lemmatization, which helps simplify the language and reduce the complexity of the analysis.

In any NLP algorithm *(fig 7)*, it is essential to ensure that all tokenized words are transformed into their root form. In the English language, a word can have multiple forms, but the aim is to treat all forms as equivalent, rather than separate meanings. This can be achieved through stemming or lemmatization. Stemming involves removing preceding letters from the word to reduce it to its base form, but it can produce incorrect spellings or entirely different meanings. In contrast, lemmatization considers the meaning of the word, making it a more accurate method. Although this step typically occurs during the preprocessing stage, perform it at this stage to ensure that the word’s part of speech is retained.

In NLP methods like text categorization, sentiment analysis, and knowledge representation, lemmatization is a crucial step. To make sure that each token is given the appropriate POS, it is crucial to carry out POS tagging following lemmatization *(Fig 8)*. Lemmatization has a significant positive impact on NLP approaches like text categorization, sentiment classification, and knowledge representation. It is possible to limit the size of the feature space, prevent overfitting, and improve the performance of machine learning models by reducing the proportion of distinctive words in the text.

E. Knowledge Extraction

The first step in creating a use-case diagram is identifying the participants, use cases, and linkages between them. We separate the input into tokens and assign the correct portion

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:42:22 UTC from IEEE Xplore. Restrictions apply.

```
def lemmatization(doc):

    doc = segmentation(doc)
    nlp = spacy.load('en_core_web_sm',disable=['ner','textcat'])
    lemmatized_text = []
    for line in doc:
        sub = []
        t = nlp(line)
        for token in t:
            if token.pos_ == 'PUNCT':
                continue
            sub.append(token.lemma_)
        lemmatized_text.append(sub)
    return lemmatized_text
```

Fig. 7. Code Snippet of Lemmatization

Port Lemmatization

User, A, and, User, B, can, equally, connect, to, WiFi, direct, and, search, available, user,
User, A, can, send, invitation, to, connect, and, user, B, can, grant, invitation,
also, they, can, send, multimedia, and, receive, multimedia, such, as, message, audio, video, and, file,
User, A, be, same, as, User, 8, can, terminate, connection,

Fig. 8. Output after Lemmatization

of the speech to each token. The actors are then located by traversing the full array in accordance with a set of rules created specifically for this task. The token can be classed as an actor if it is a subject, noun, or pronoun and its dependency is compound. We search the code for tokens that contain both verbs and objects to find use cases, eliminating helping verbs because they are different from the primary verbs [6]. This is how the logic is implemented *(fig 9)*:

The first one needs to be a verb, and the second one ought to constitute a noun, according to the use case's main rule [12]. Handling exceptions is also a crucial task as it can emerge when writing two verbs together like scroll up/scroll down content which can be handled by segregating verbs and adding a noun to each of them. An actor and a use case diagram have a relationship that is established by identifying the tags of the GKPs, if the actor and the use case are in the same sentence boundary then they have a relationship. After the last step, have knowledge about all the actors, and their respective use cases *(fig 10)*. Now to give the use case diagram life, use this information gain and convert all the findings into images using PIL.

```
while c < len(pos_text):
    actors = set()
    while c < len(pos_text):
        if (pos_text[c].pos_ == 'NOUN' or
        pos_text[c].pos_ == 'PROPN') and
        pos_text[c].dep_ == 'compound':
            if pos_text[c+1].text not in names:
                actors.add((pos_text[c].lemma_ + " " + pos_text[c+1].lemma_).title())
            c += 1
        elif (pos_text[c].pos_ == 'NOUN' or pos_text[c].pos_ == 'PROPN'):
            if pos_text[c].text not in names:
                actors.add(pos_text[c].lemma_.title())
        elif pos_text[c].pos_ == 'PRON':
            actors = actors.union(prev_actors)
        elif pos_text[c].pos_ == 'VERB':
            break
        c += 1
```

Fig. 9. Code Snippet of Knowledge Extraction

Knowledge Extraction

User A: send invitation invitation, search user, connect, receive multimedia, terminate connection, send multimedia,
User B: connect, search user,
User: terminate connection,

Fig. 10. Output of Knowledge Extraction

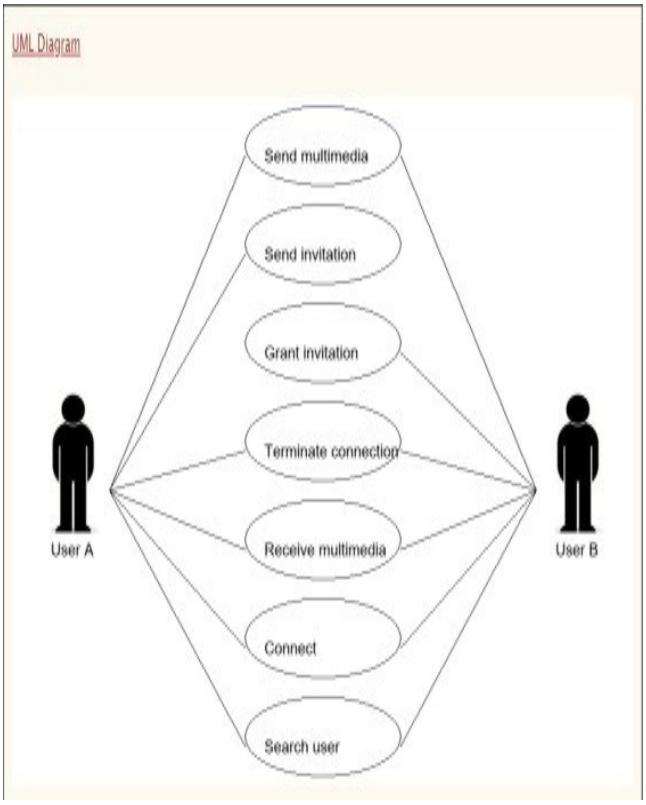


Fig. 11. Result

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:42:22 UTC from IEEE Xplore. Restrictions apply.

VI. CONCLUSION

In conclusion, this research study has focused on analyzing the generation of UML diagrams in the context of software engineering. In order to ensure that the project developers have the same knowledge and to clear up any misunderstandings during the software development process, UML diagrams are essential. The proposed study is focused on creating a use case model from the functional requirements of SRS, which results in a streamlined form of the original problem statement.

In order to address this issue, this study has created an approach by using natural language processing to examine and extract knowledge from given requirements. We were able to successfully gather the data required to produce use case diagrams through segmentation, tokenization, part- of-speech tagging, lemmatization, and knowledge extraction. Notably, when producing these diagrams, our technique has shown an accuracy rate of less than a fourth. We have built a portal where users may access the gathered data and inspect the resultant UML diagrams in order to promote cooperation and simplify the software development process. With this endeavor, the engineering team will be more effective and cohesive and be able to go forward with the creative process as a unit. The proposed strategy preserves the organic essence of the industry while maintaining a uniform and thorough under- standing of the functional needs by utilizing natural language processing tools. The discipline of software engineering and its practitioners stand to gain from additional developments in automating UML diagram production as a result of this research.

VII. FUTURE SCOPE

Based on the current state of the research topic, there are several potential areas for future exploration and development. Here are some possible directions for future research:

- Improving accuracy and efficiency: The effectiveness and validity of the process used to create the use case model from the functional requirements of the SRS document must be investigated. This can entail looking into new techniques and algorithms for discovering relevant information and automating the diagram-generating process.
- Extending to other types of diagrams: The approach could be expanded to create other sorts of UML diagrams, however, this research concentrates on creating use case diagrams. This can entail changing the algorithm to recognize and retrieve pertinent data.
- Evaluation and comparison: Future studies should compare the accuracy, efficiency, and effectiveness of the suggested approach to other approaches for creating use case diagrams using real-world datasets.
- Integration with software development tools: By providing automatic support for creating and updating UML diagrams, integrating the suggested technique with current software development instruments and platforms could speed up the software development process. Overall, there are many potential directions for future research in this area, and the continued exploration and development of these ideas could lead to significant advances in the field of software engineering.

REFERENCES

[1]H Saiedian, R Dale, Requirements engineering: making the connection between the software developer and customer, Information and Software Technology, Volume 42, Issue 6,2000, Pages 419-428, ISSN 0950-5849,https://doi.org/10.1016/S0950-5849(99)00101-9.

[2]W. J. Lloyd, M. B. Rosson, and J. D. Arthur, "Effectiveness of elicitation techniques in distributed requirements engineering," Proceedings IEEE Joint International Conference on Requirements Engineering, Essen, Germany, 2002, pp. 311-318, doi: 10.1109/ICRE.2002.1048544.

[3]E. A. Abdelnabi, A. M. Maatuk, and M. Hagal, "Generating UML Class Diagram from Natural Language Requirements: A Survey of Approaches and Techniques, " 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, 2021, pp. 288-293, doi: 10.1109/MI-STA52233.2021.9464433.

[4]S. Falkenstine, A. Thornton and B. Meiners, "Natural Language Processing for Autonomous Identification of Impactful Changes to Specification Documents," 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2020, pp. 1-9, doi: 10.1109/DASC50938.2020.9256611

[5]Alksasbeh, Malek and Alqaralleh, Bassam and Alramadin, Tahseen and Alemerien, Khalid. (2017). An Automated Use Case Diagrams Generator From Natural Language Requirements. Journal of Theoretical and Applied Information Technology. 95. 1182-1190.

[6]Z. A. Hamza and M. Hammad, "Generating UML Use Case Models from Software Requirements Using Natural Language Processing, " 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO), 2019, pp. 1-6, doi: 10.1109/ICMSAO.2019.8880431.

[7] Martijn B. J. Schouten, Guus J. Ramackers, Suzan Verberne, "Preprocessing Requirements Documents for Automatic UML Modelling", Natural Language Processing and Information Systems, vol.13286, pp.184, 2022.

[8] Fatma Bozyi ğit, Özlem Aktaş, Deniz Kılınç, "Linking software requirements and conceptual models: A systematic literature review", Engineering Science and Technology, an International Journal, vol.24, no.1, pp.71, 2021.

[9] Sharif Ahmed, Arif Ahmed, Nasir U. Eisty, "Automatic Transformation of Natural to Unified Modeling Language: A Systematic Review", 2022 IEEE/ACIS 20th International Conference on Software Engineering Research, Management, and Applications (SERA), pp.112-119, 2022.23

[10] Esra A. Abdelnabi, Abdelsalam M. Maatuk, Tawfig M. Abdelaziz, "An Algorithmic Approach for Generating Behavioral UML Models Using Natural Language Processing", The 7th International Conference on Engineering and MIS 2021, pp.1, 2021.

[11]S. Vemuri, S. Chala and M. Fathi, "Automated use case diagram generation from textual user requirement documents, " 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017, pp. 1-4, doi:10.1109/CCECE.2017.7946792.

[12]S Aithal, P Desai and S Vinay, "An Approach towards Automation of Requirements Analysis", Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, 2009.

[13]S Kotsiantis, V Tampakas and M Ikonomakis, "Text Classification Using Machine Learning Techniques", WSEAS TRANSACTIONS on COMPUTERS, vol. 4, no. 8, pp. 966-974, August 2005.

[14]S Aithal, P Desai S Vinay, "An Approach towards Automation of Requirements Analysis, " in Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, Hong Kong, 2009.

[15]R Garigliano L Mich, "NL-OOPS: A Requirements Analysis tool based on Natural Language Processing, " in Conference on Data Mining, 2002, Vol. 3.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:42:22 UTC from IEEE Xplore. Restrictions apply.

[16]H. M. Harmain and R. Gaizauskas, "CM-Builder: an automated NL-based CASE tool," Proceedings ASE 2000. Fifteenth IEEE International Conference on Automated Software Engineering, Grenoble, France, 2000, pp. 45-53, doi: 10.1109/ASE.2000.873649.

[17]S Kotsiantis, V Tampakas M Ikonomakis, "Text Classification Using Machine Learning Techniques, " WSEAS TRANSACTIONS on COM- PUTERS, vol. 4, no. 8, pp. 966-974, August 2005

[18] Kadir Yalcin, Ilyas Cicekli, Gonenc Ercan, An external plagiarism detection system based on part-of-speech (POS) tag n-grams and word embedding, Expert Systems with Applications, Volume 197, 2022, 116677, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2022.116677.

[19] Julian Kupiec, Robust part-of-speech tagging using a hidden Markov model, Computer Speech and Language, Volume 6, Issue 3,1992, Pages 225-242, ISSN 0885-2308,https://doi.org/10.1016/0885-2308(92)90019- Z.

[20] Btoush, Eman S., and Mustafa M. Hammad. "Generating ER diagrams from requirement specifications based on natural language processing." International Journal of Database Theory and Application 8.2 (2015): 61-70.

[21] Mohammed, A.H.K., Jebamikyous, HH., Nawara, D. et al. IoT text analytics in the smart education and beyond. J Comput High Educ 33, 779–806 (2021). https://doi.org/10.1007/s12528-021-09295-x.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 02,2023 at 06:42:22 UTC from IEEE Xplore. Restrictions apply.

Wireless Personal Communications
https://doi.org/10.1007/s11277-023-10680-1

Network Traffic Analysis in Software-Defined Networking Using RYU Controller

Shanu Bhardwaj[1] · **Ashish Girdhar**[1]

Accepted: 12 July 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Software-Defined Networking (SDN) has emerged as a promising paradigm to enhance network control and management by decoupling the planes. With SDN, the centralized controller plays a critical role in managing network resources and traffic flows. Throughout the most recent couple of years, networks turned out to be more imaginative for developing different applications with the help of SDN. Network traffic analysis is a vital task in understanding network behaviour, identifying anomalies, and optimizing network performance. To deal with the load of changes in the networking industry, there is an extraordinary requirement for a productive SDN controller to work on the usage of network resources for a better presentation of the network. Therefore, the proposed approach leverages the RYU controller, an open-source SDN controller framework, to collect and analyse network traffic data. By utilizing RYU's capabilities, we can dynamically monitor and capture network traffic statistics, such as bandwidth, throughput, packet counts, and Round trip time (RTT). These statistics provide valuable insights into network performance, and traffic patterns. By leveraging real-time traffic analysis, we can dynamically adjust routing paths, and allocate network resources efficiently. Hence, the proposed work assesses the development of SDN architecture through a network topology and then, implementation of RYU controller has been done to evaluate various network performance parameters. To evaluate the effectiveness of our approach, we conduct experiments using a simulated SDN environment. We compare the performance parameters of our traffic analysis techniques with traditional methods and showcase the advantages of utilizing SDN and the Ryu controller for network traffic analysis. The results demonstrate that our approach provides accurate and timely insights into network traffic behaviour, facilitating efficient network management. In conclusion, this study highlights the significance of network traffic analysis in SDN environments and demonstrates the effectiveness of the Ryu controller for extracting valuable insights from network traffic data.

Keywords Software-defined networking · Wireless networks · Network topology · Traffic network · Traffic engineering · SDN controller · Internet of Things

✉ Shanu Bhardwaj
shanubhardwaj1@gmail.com

[1] Department of Computer Science and Engineering, Delhi Technological University, Delhi, India

Published online: 30 July 2023

1 Introduction

These days, the world behaves like a universal home where the different sections of this world are collaborated by telecommunications, named the Internet. Today, the number of Internet users reached up to three billion [1]. To inculcate all these users under a roof, it is necessary to change the current scenario of the network architecture. There is a need to interconnect all things under the concept of IoT [2] and cloud computing [3]. In the traditional architectural network [4], the router contains both planes [5]. The control plane has a routing algorithm in the router while the data plane, which is also known as a forwarding plane [6] used to forward the data packets to the next-hop router [7]. But in the emerging paradigm, both the planes are separated to show their behaviour uniquely and the paradigm is termed as SDN.

SDN is a progressive networking system that upholds heterogeneous organizations with fast development and dynamic by utilizing diverse programmable planes [9]. The requirement for this organization is to help the availability of zillions of contraptions and speedy specialist co-ops to course the traffic [10]. Also, a domain-explicit language, P4 satisfies these prerequisites and has known wide reception over the years from the networking industry and academic field [11]. It is an unmistakable systems paradigm that hopes to change the boundaries of conventional organization foundations by moving all the control functionalities to a brought-together region [12] as shown in Fig. 1. The SDN architecture is having three layers as follows, also depicted in Fig. 2:

- Physical layer: It is the first layer of the architecture. This layer consists of various networking devices that are connected to communicate the data traffic [13].
- Control layer: The control layer is the middle layer that is responsible for managing the whole network. This layer consists of the brain of the network which is known as the controller. A controller is the central single entity that tracks traffic moving in the network. It provides the best route to the data to make the network more effective by eliminating the waiting time of the network. [14].
- Management layer: The management layer is the uppermost layer of the SDN environment which is responsible for keeping records of data traffic by managing documentation, file database, and other applications [15].

SDN permits the detachment of information and control plane, therefore, switches will turn into the sending gadgets. In SDN, regulators are the mind of the organization which is having the force of performing control choices during the directing of information bundles [16]. SDN rose as a method to extend the convenience of the organization, decreasing expenses, lessening the intricacy of equipment, and permitting creative examination [17]. The architecture of SDN consists of three different layers named physical, network, and application layers [18]. The physical layer has also termed an infrastructure or data plane in which the hub of network devices, Open Flow switches [19], and other infrastructure are connected. The network layer consists of a controller having different libraries and protocols supported under the norms of the controller [20]. The application layer is having various user applications, controller applications, operators, and open stacks for end-to-end data transmission [21]. Various application-based routing protocols have been developed in recent years for WSNs such as energy efficiency [22, 23], resource management, traffic analysis, and many more. The paper

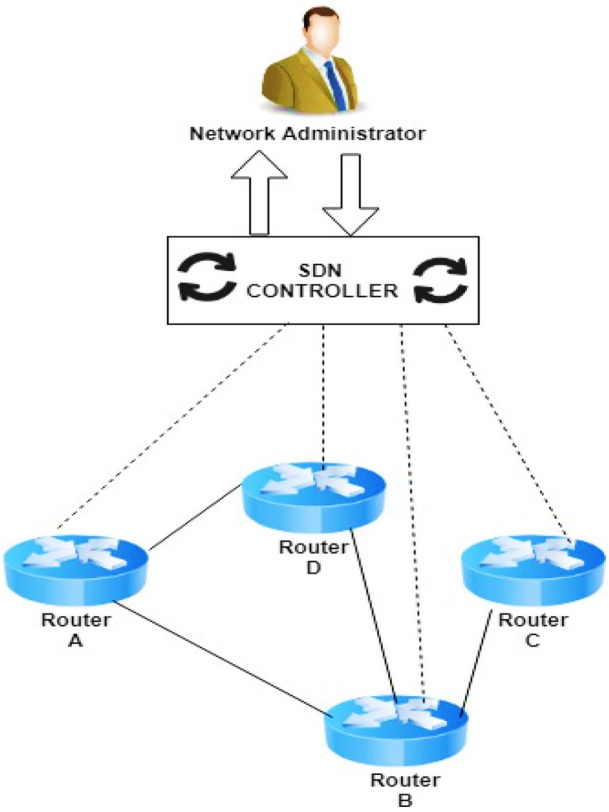


Fig. 1 Overview of SDN [8]

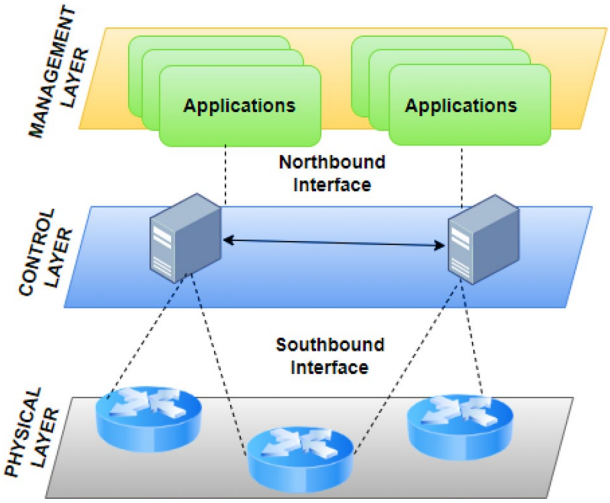


Fig. 2 Three-layered architecture of SDN

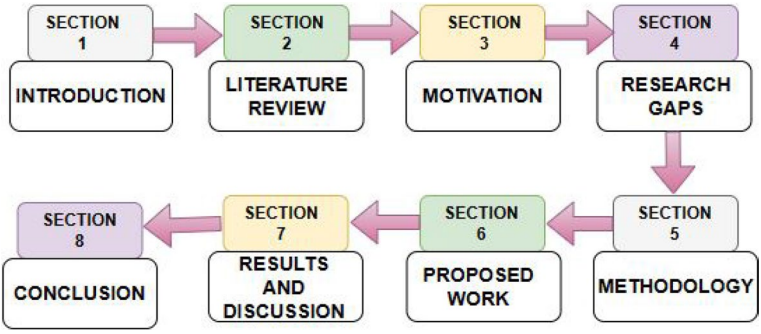


Fig. 3 Structure of the paper

also provides the framework for traffic analysis in the field of SDN using network topology having an RYU controller implemented.

The major contributions of the paper are as follows:

- Established a network custom topology.
- Integrated the Open flow RYU controller into the network topology.
- Evaluated the performance of the RYU controller.
  - Mininet tool and Wireshark traffic analyzer tool are used to create the setup environment.
  - Hping3 and IPERF tools are used to get the performance parameters.
- Presented a novel framework to analyse the network data traffic based on the RYU controller.

The rest of the research paper is structured in the following manner as shown in Fig. 3.

2 Literature Review

There is a summarization of the existing literature which has already been done in SDN. The main task of reviewing the previous work is to gather the objective of the research available on the application of SDN and its components by evaluating results as per the study approaches shown in Table 1.

After conducting a literature review several key findings have emerged:

- The literature features that data traffic analysis in the network utilizing the Ryu SDN controller offers a few advantages. It empowers real-time checking and investigation of data traffic, permitting network heads to identify and answer anomalies or security breaks instantly. It additionally works with network enhancement and resource assignment based on the design of traffic, further development improves the performance of the network.
- Furthermore, the literature highlighted that the combination of the functionalities of SDN with traffic analysis, like TE or QoS, can upgrade the network productivity as

Table 1 Comparative study of traffic analysis in SDN

Reference no	Year	Authors	Techniques/tools	Summary
[24]	2013	J. Chica et al	Software-Defined Networking, Cbench	• The paper reviewed the performance examination of publically accessible OpenFlow regulators like Beacon, NOX, Maestro, and NOX-MT • Designed a series of flow-based benchmarks
[25]	2013	M. Caria et al	TA2 network, SNDlib library	• The paper proposed an algorithm for the optimal substitution of the hosts in the network. Hence, it reduces the capacity of the network to enhance the TE
[26]	2016	I. Akyildiz et al	Software-Defined Networking	• The paper investigated the solution for SDN traffic • Also, presented the challenges of SDN traffic engineering on some key parameters namely, accuracy, availability, and scalability
[27]	2017	M. Wang et al	Mininet	• Proposed an answer for the solid transmission of information bundles starting with one host and then onto the next to have using the SDN climate • The proposed structure gave an all-around directed utilization of data transfer capacity just as the dependable solid parcel transmission
[28]	2017	L. Yu et al	Peering BGP testbed	• In the proposed SDN architecture, the SDN controller allowed the user to manipulate the IP • The paper concluded that in the future, the security analysis factor could be added to enhance the capability of a proposed SDN architecture
[29]	2017	I. Bholebawa et al	Mininet, OpenvSwitch	• The paper gave the presentation examination of two SDN regulators to be specific, Floodlight and POX. Distinctive SDN geographies were intended for the two regulators and thought about the exhibition utilizing the boundaries, throughput, and full-circle time • As per the presentation, the Floodlight regulator gave more powerful outcomes when contrasted with POX
[30]	2017	Y. Guo et al	Online Offline Routing Optimizer (OORO)	• The paper first formulated the issues related to traffic engineering and then try to sort the network using the proposed algorithm
[31]	2018	R. Amin et al	Mininet, EstiNet, GNS3	• The paper presented a comprehensive survey in the field of a hybrid SDN network. Also, identified the research gaps in the existing solutions
[32]	2018	S. Singh et al	-	• The paper compared the performance analysis of different algorithms that were operated in the SDN control plane. The parameters taken to perform the comparison were jitter, latency, Q-factor, and many more

Table 1 (continued)

Reference no	Year	Authors	Techniques/tools	Summary
[33]	2019	S. Mehr et al	SVM	• Proposed a flow to detect an attack on the RYU SDN controller. The author used machine learning to detect the attack using the SVM environment • Also, claimed that the results showed a 36% of reduction in attacks using the proposed method
[34]	2019	J. Bhatia et al	Tensorflow Keras	• Proposed a calculation for the continuous information transmission directed under the SDN controller to assess the various boundaries, for example, thickness and clog forecast of information traffic
[35]	2019	A. V. Priya et al	D-ITG	• The creator thought about the presentation boundaries of OpenFlow regulators named RYU, FloodLight, and POX. The examination was done dependent on the bundle taking care of the limit on the D-ITG apparatus • The creator inferred that the FloodLight regulator shows a superior show as far as throughput and postponement when contrasted with that of others • The creator proposed that later on the work can be additionally stretched out by looking at the OpenDayLight and OpenContrail regulators
[36]	2019	S. Badotra et al	Mininet, Wireshark	• The author compared two open-source controllers of SDN named OpenDayLight and Open Networking operating systems
[37]	2019	W. Queiroz et al	Mininet, Iperf3	• The proposed approach defined the usage of network bandwidth in the SDN and suggested that in the future the same proposed method can be effectively implemented by decreasing the interval time so that the performance would increase
[38]	2020	M. Islam et al	Mininet	• The paper implemented SDN architecture to perform the analysis of the traffic on the RYU SDN controller • Also, fascinated the researchers to perform this RYU controller implementation using different topologies in the future to evaluate the performance
[39]	2021	Kurungadan et al	No simulator was used	• The paper utilizes SDN to control the heap of the WiFi traffic. The gateways get information from IoT gadgets using remote connections • The author proposed three SDN algorithms that are conceived to diminish the delay in packet forwarding
[40]	2022	Bhardwaj et al	Mininet	• The author analyzed and monitored the real-time data traffic via implementing the architecture of SDN and also examined the performance of the controller in the proposed topology using some of the network performance parameters

Table 1 (continued)

Reference no	Year	Authors	Techniques/tools	Summary
[41]	2023	Raja et al	Deep Learning models	• The author provided an empirical study on the flow of data traffic in the network using the prediction method in SDN • Also, the paper presented the data flow technique based on different approaches such as flow count, entropy technique as well as DL approach
[42]	2023	P. Shukla et al	MATLAB	• The recommended strategy is centered around showing the flow of data traffic and also protecting the data forwarding layer • The proposed model accomplishes a noteworthy performance of attack detection in DDoS assaults by forestalling network-sending execution degradation

well as the client experience. The capacity to reconfigure the data paths and strategies dynamically, given traffic analysis is an important component for network administrators.

3 Motivation

To the finest of our ability, there is not any research that provides the traffic analysis in the RYU SDN controller for traffic engineering to evaluate the network parameters.

The motivation is to execute the design of SDN utilizing the RYU SDN controller for the data traffic investigation. In systems administration enterprises, scholastic fields, and server farms, there is an extraordinary requirement for an elite controller. Subsequently, it is significant to research the performance of an open-source controller of SDN [38]. The execution of the SDN engineering is to be completed in the Mininet emulator incorporating the RYU controller with the part of exchanging center, four hubs consider under solitary topology, and an OpenFlow switch. The work intends to assess execution boundaries like transmission and receiving of packets as well as bytes, and time is taken depending upon the experimental conditions. The significance of the research study is also depicted in Table 2.

4 Research Gaps

Many researchers have given their best in the field of SDN to enhance the network management approach. But still, many challenges are faced in the SDN environment to generate a smooth flow of data traffic. Hence, a lot of effort can be done to overcome these challenges. Based on those challenges, some of the major research gaps have been identified:

- The current research on network traffic analysis using the Ryu controller often focuses on static traffic patterns or predefined rules. Investigating dynamic traffic engineering approaches that leverage traffic analysis insights to optimize network performance and resource allocation in real time is an area that requires further exploration [10].

Table 2 Significance of Research Study

Significance of proposed research	• There is an extraordinary requirement for a superior controller in networking sectors, academic centers, data fields, and so on • The exhibition of the SDN regulator gives a brilliant effect on intensifying the capacities of the SDN organization. Hence, it is essential to explore the presentation of the SDN controller
Spots of the problem	• The current methodologies are short falling as far as giving the traffic evaluation of a particular SDN controller • An adequate examination is absent in the space of the traffic execution assessment of the RYU SDN regulator
Results Reported	• The exploration work will be adequate for the analysts, industrialists, and systems administration skills to pick the RYU controller for future organizations, research applications, and module advancements

- There is no such specific study on the RYU SDN controller which provides the performance characteristics of custom topology [38].
- The existing approaches are lacking in terms of providing the traffic performance evaluation of a specific SDN controller instead of comparison [35, 39, 43].
- There is very little work in which those controllers are chosen that are scripted in emerging languages such as Python for traffic analysis in SDN environment [26].

5 Methodology

Three segments are expected to execute the SDN architecture. The first segment is SDN Applications in which different programs convey practices and required resources with the controller of the SDN through APIs. The other segment is the RYU SDN controller which is a consistent element that gets instructions or prerequisites from the SDN Application layer and transfers them to the system's networking element. The last segment is the Networking elements of the SDN structure, which control the sending and information-handling abilities of the SDN network. This segment incorporates the sending and preparing of the information way. In real-time, establishing an actual framework of SDN is a very tedious task. To tackle this issue, one can utilize a specific test emulator to establish the connection in the SDN environment. There are a few devices utilized as a simulator, for example, OMNET++ [44], OFNet [45], NS-3 [46], Mininet [47], and so on. In the literature section, work done on mostly used testbeds used in the experimental setup has been evaluated. Among them, Mininet appears to function admirably as shown in Fig. 4. Mininet is considered as a popular network emulator that allows the creation of virtual networks for testing and development purposes and also. This examination has played out a performance assessment through the RYU SDN controller utilizing several tools such as iperf3 [48], ping, and others too.

Mininet is created to help exploration in SDN. It is an extraordinary open-source emulator. Mininet seems to be friendly as it uses the interface of the command line. The explanation for this is that Mininet utilizes the namespaces of the network because of its virtualization innovation, it can uphold numerous virtual hosts without hindering as well as slowing the system. Mininet creates a realistic virtual organization with

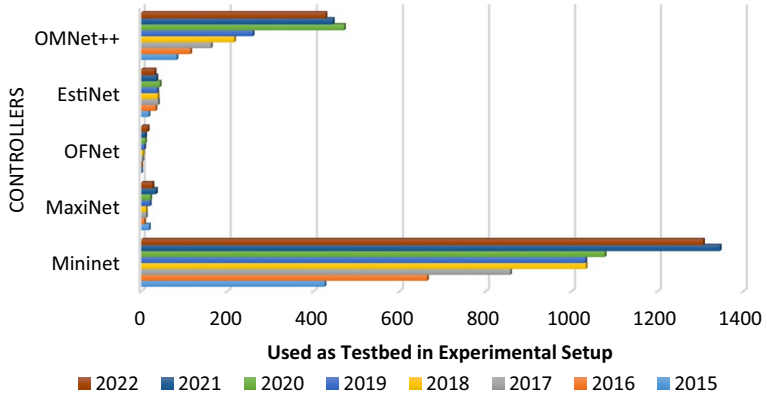


Fig. 4 Work done on different emulators used in SDN

switches, nodes, and several application segments, running kernels all at one machine. By this means, the virtual scenario of SDN can be fully designed with the help of a Mininet emulator, which comprises an SDN controller. Mininet also upholds a few inherent essential network topologies such as (1) Single topology, (2) Minimal topology, (3) Linear topology, (4) Tree topology, and so forth. For the execution of these topologies, the Mininet interface can be used by executing a solitary line order.

6 Proposed Work

Implementing the physical framework of an SDN architecture in the real world is a very expensive and difficult task which is the reason, there are many tools to create an SDN environment used as a simulator or emulator. In the implementation, first, we need to create an SDN architecture to transfer the data traffic from one node to a different node via a specific SDN controller in the SDN emulator known as Mininet. The SDN architecture is going to be implemented using three constituents: Network devices, controllers, and applications as depicted in Fig. 5. Then, the emulation and evaluation of the proposed architecture will be done based on various performance parameters as depicted in Fig. 6.

The flow of the research is categorized into five segments mentioned below:

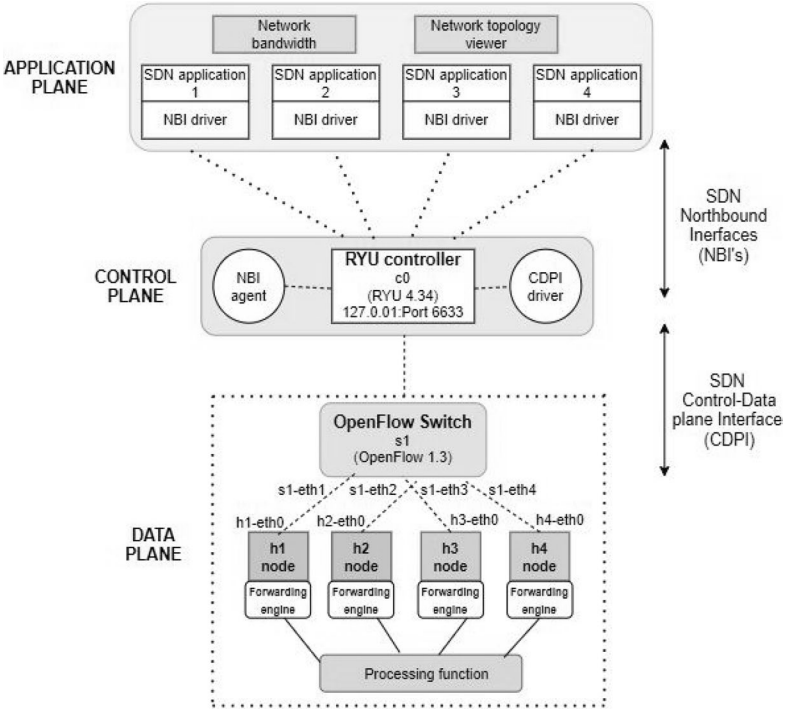


Fig. 5 Components of the proposed research work

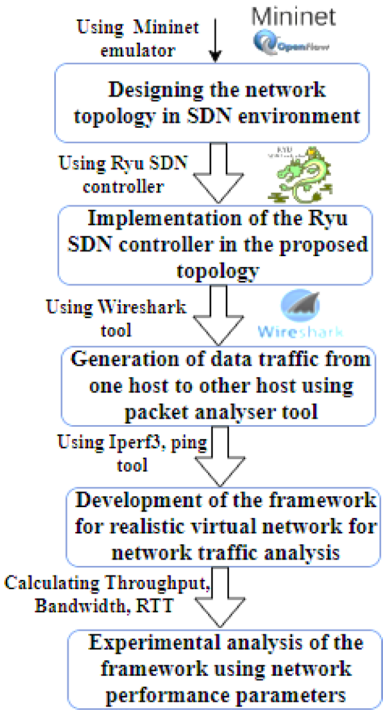


Fig. 6 Flow of the proposed research work

1. Designing the network topology
2. Implementation of Ryu controller in the designed topology
3. Generation of data traffic in the network
4. Development of the framework for a realistic virtual network
5. Experimental analysis using network performance parameters.

The expected outcome from the emulation spotlights the performance criteria of the RYU SDN controller. Hence, the proposed research work will be sufficient for the researchers, industrialists, and networking experts to choose the RYU controller for future businesses, research applications, and module developments. The below-mentioned Fig. 5 describes the components of the proposed work:

6.1 Setup Environment

A custom topology made in the Mininet apparatus has four hubs and an OpenFlow switch. The connections are made to make the association among the hosts and switch, for example, (h1, s1), (h2, s1), and so on. After the actual switch creation, it would involve the capacities of the comparing controller, as displayed in Fig. 7. By this progression, the OpenFlow switch 1.3 was stacked to communicate the information parcels starting with one host and then onto the next. The information parcels were produced in the organization with the assistance of the Wireshark tool. These information bundles were sent from the source host to the objective host.

```
mininet> h1 ping -c 5 h2
PING 10.0.0.2 (10.0.0.2) 56(84) bytes of data.
64 bytes from 10.0.0.2: icmp_seq=1 ttl=64 time=33.3 ms
64 bytes from 10.0.0.2: icmp_seq=2 ttl=64 time=0.214 ms
64 bytes from 10.0.0.2: icmp_seq=3 ttl=64 time=0.070 ms
64 bytes from 10.0.0.2: icmp_seq=4 ttl=64 time=0.057 ms
64 bytes from 10.0.0.2: icmp_seq=5 ttl=64 time=0.059 ms

--- 10.0.0.2 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4069ms
rtt min/avg/max/mdev = 0.057/6.745/33.326/13.290 ms
```

Fig. 7 Transmission of data packets in the proposed topology

```
--- 10.0.0.2 ping statistics ---
5 packets transmitted, 5 received, 0% packet loss, time 4078ms
rtt min/avg/max/mdev = 0.058/4.769/23.453/9.342 ms
mininet> links
h1-eth0<->s1-eth1 (OK OK)
h2-eth0<->s1-eth2 (OK OK)
h3-eth0<->s1-eth3 (OK OK)
h4-eth0<->s1-eth4 (OK OK)
mininet> ports
s1 lo:0 s1-eth1:1 s1-eth2:2 s1-eth3:3 s1-eth4:4
mininet> h4 iperf -s &
mininet> h1 iperf -c h4
------------------------------------------------------------
Client connecting to 10.0.0.4, TCP port 5001
TCP window size: 85.3 KByte (default)
------------------------------------------------------------
[  3] local 10.0.0.1 port 42134 connected with 10.0.0.4 port 5001
[ ID] Interval       Transfer     Bandwidth
[  3]  0.0-10.0 sec  4.78 GBytes  4.11 Gbits/sec
mininet>
```

Fig. 8 Connected ports and links in the proposed topology

Figure 8 depicted the successful transmission of the data from the source host to the destination host with 0% packet loss. Also, the link association between each node to every other node is successfully created. Each node is further associated with different ports and the ports are then connected to the main controller.

7 Results and Discussion

The segment reports the outcomes acquired for the SDN network execution assessment with RYU. The presentation boundaries like bandwidth, throughput, and RTT have been assessed under TCP, and UDP information traffic utilizing a few benchmarking apparatuses.

Figure 9 demonstrated the source and the destination port connection having a TCP window size of 85.3 Kbyte. The output of the screen is categorized into three parameters: Time interval, Transfer rate, and Bandwidth as depicted in Fig. 10. The time interval taken

```
mininet> h1 iperf -c h4 -d
------------------------------------------------------------
Server listening on TCP port 5001
TCP window size: 85.3 KByte (default)
------------------------------------------------------------
------------------------------------------------------------
Client connecting to 10.0.0.4, TCP port 5001
TCP window size:  298 KByte (default)
------------------------------------------------------------
[  3] local 10.0.0.1 port 53500 connected with 10.0.0.4 port 5001
[  5] local 10.0.0.1 port 5001 connected with 10.0.0.4 port 33086
[ ID] Interval       Transfer     Bandwidth
[  3]  0.0-10.0 sec  4.07 GBytes  3.50 Gbits/sec
[  5]  0.0-10.0 sec  1.46 GBytes  1.25 Gbits/sec
```

Fig. 9 Connections among the ports

```
mininet> h1 iperf -c h4 -P 5
------------------------------------------------------------
Client connecting to 10.0.0.4, TCP port 5001
TCP window size: 85.3 KByte (default)
------------------------------------------------------------
[  7] local 10.0.0.1 port 53514 connected with 10.0.0.4 port 5001
[  3] local 10.0.0.1 port 53506 connected with 10.0.0.4 port 5001
[  4] local 10.0.0.1 port 53508 connected with 10.0.0.4 port 5001
[  6] local 10.0.0.1 port 53512 connected with 10.0.0.4 port 5001
[  5] local 10.0.0.1 port 53510 connected with 10.0.0.4 port 5001
[ ID] Interval       Transfer     Bandwidth
[  7]  0.0-10.0 sec  1.39 GBytes  1.19 Gbits/sec
[  3]  0.0-10.0 sec  1.43 GBytes  1.22 Gbits/sec
[  4]  0.0-10.0 sec  1.43 GBytes  1.23 Gbits/sec
[  6]  0.0-10.0 sec  1.43 GBytes  1.22 Gbits/sec
[  5]  0.0-10.0 sec  1.41 GBytes  1.21 Gbits/sec
[SUM]  0.0-10.0 sec  7.09 GBytes  6.07 Gbits/sec
```

Fig. 10 TCP window size of the network

Table 3 The bandwidth of the proposed topology

Time interval (s)	Transfer (GBytes)	Bandwidth (Gbps)
0–10	1.39	1.19
0–10	1.43	1.22
0–10	1.43	1.23
0–10	1.43	1.22
0–10	1.41	1.21
Total time taken	**7.09**	**6.07**

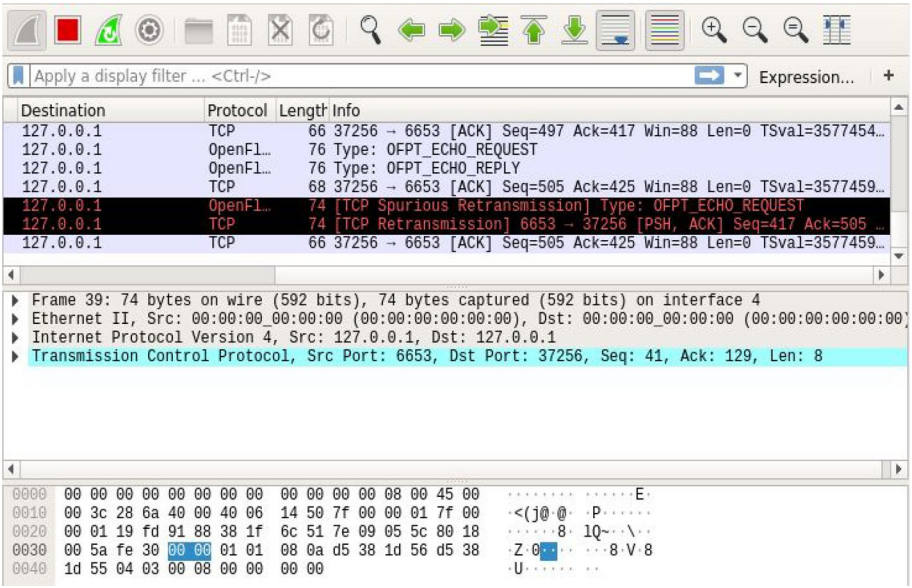


Fig. 11 Screenshot of the packet analysis using the Wireshark tool

here is between 1 to 10 seconds. Table 3 details the experimental values of the connected ports. Figure 11 depicted the flow of the data packets analyzed using a packet analyzer tool named Wireshark tool. The data packets used in the network are generated from the packet analyzer tool.

7.1 Bandwidth

Bandwidth in networking refers to the maximum amount of data that can be transmitted over a network connection in a given period. It represents the capacity or throughput of the network and is typically measured in bits per second (bps). It is influenced by various factors such as the physical properties of the network medium, the network infrastructure, and the technologies used for data transmission.

$$\text{Bandwidth(in bps)} = \frac{Data\ rate\ (in\ bits)}{Time\ (in\ seconds)} \tag{1}$$

Bandwidth is a critical component in determining the maximum data transfer capacity of a network connection. It directly affects the rate at which data can be transmitted and received, impacting the overall performance and responsiveness of the network. Hence, the bandwidth of the proposed topology is elaborated in Table 3. The experimental output is categorized into three parameters: Time interval, Transfer rate, and Bandwidth. The time interval taken here is between 1 to 10 seconds. Table 3 details the experimental values of the connected ports of the proposed setup.

7.1.1 Comparison and Analysis of Proposed Bandwidth with Classical Bandwidth

The proposed network topology offers a significant increase in bandwidth compared to the existing setup. By implementing a custom topology, where each device is connected to multiple other devices, the network can leverage parallel data paths and distribute traffic more efficiently. This distribution of connections helps alleviate bottlenecks and congestion points that may occur in classical topology.

In the case of the proposed topology, if it has been designed with bandwidth optimization in mind, it aims to increase the available capacity for data transmission. The proposed topology may prioritize efficient utilization of available bandwidth, potentially resulting in a higher overall network capacity as shown in Table 4.

7.2 Throughput

Throughput refers to the rate at which data can be successfully transmitted or processed through a system, network, or device. It is typically measured in terms of data transfer rate, often expressed in bits per second (bps).

In networking, throughput represents the amount of data that can be transmitted over a network connection within a specific time frame. It is influenced by several factors, including bandwidth, latency, packet loss, and network congestion. Higher bandwidth allows for greater throughput as it provides a larger capacity for data transmission. Lower latency and minimal packet loss also contribute to improved throughput by reducing delays and ensuring reliable delivery of data packets. Throughput addresses the greatest measure of information traffic handled using regulators between two hubs of the organization in a second. In TCP traffic, one of the hosts goes about as a customer, and one more goes about as a server. iperf3 utility has been used to test the throughput of the regulator. On the client side, iperf3 has been executed in 10 seconds, and the data about throughput has been assembled on one more side of the organization. The aftereffect of execution is displayed and shown graphically in Fig. 10.

$$\text{Throughput} = \frac{\text{Amount of data received}}{\text{Given time frame}} \tag{2}$$

Table 4 Bandwidth comparison of the proposed and classical topology

Nodes	Bandwidth of classical topology (Gbps)	Bandwidth of proposed topology (Gbps)
h1 to h2	25.1	26.7
h1 to h3	24.5	25.2
h2 to h3	24.4	25
h1 to h4	NA	24.8
h2 to h4	NA	23.8
h3 to h4	NA	22.7

7.2.1 Comparison and Analysis of Proposed Throughput with Classical Throughput

Throughput is a critical metric that measures the rate at which data can be successfully transmitted through a network. It depends on factors such as bandwidth, latency, packet loss, network congestion, and the efficiency of network components. In the case of the proposed topology, by incorporating higher-capacity links, efficient routing algorithms, and advanced traffic management techniques, the proposed topology can potentially enhance the overall throughput of the network. This improvement results in faster data transmission and increased capacity to handle concurrent data streams as shown in Fig. 12.

On the other hand, the previous topology has limitations in terms of available bandwidth, outdated infrastructure, or suboptimal routing that could impact throughput. It might suffer from congestion points, inefficient resource allocation, or network bottlenecks, which can hinder the data transfer rate and limit overall throughput as shown in Table 5.

The comparison of throughput between the proposed and default topology depends on the specific design choices, network infrastructure, and optimization strategies employed. The proposed topology addressed the limitations of the previous one and incorporated enhancements to maximize bandwidth, minimize latency, and improve traffic management, it has the potential to significantly increase throughput and enhance the network's overall performance.

7.3 Round Trip Time

In SDN, RTT refers to the time it takes for a packet to travel from a source to a destination and back again. SDN controllers like Ryu (as used in the article) can be used to perform traffic analysis and measure RTT in SDN networks. RTT is additionally known as ping test time. RTT is the absolute time taken to by the information bundle to reach from a particular source host to the objective host and the affirmation parcel to reach back as depicted in Fig. 13.

The round trip time is a crucial metric that measures the time taken for a data packet to travel from a source to a destination and back again. It is influenced by various elements, including network latency, packet loss, congestion, and the number of network

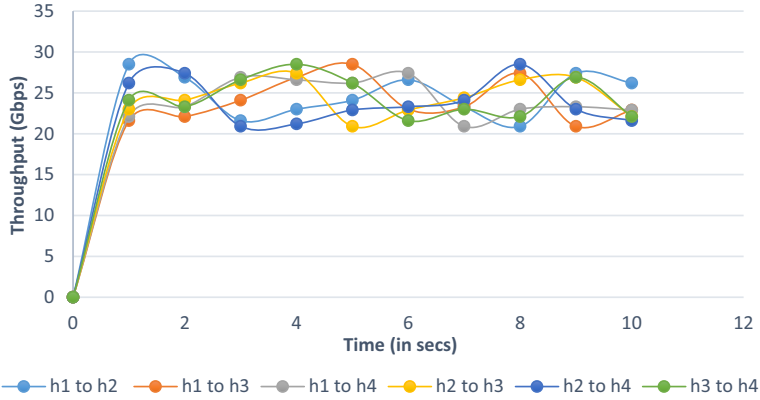


Fig. 12 TCP throughput

Table 5 Throughput comparison of the proposed and the classical topology

Nodes	Default SDN topology		Proposed SDN topology	
	Minimum throughput	Maximum throughput	Minimum throughput	Maximum throughput
h1 to h2	21.2	27.4	22.6	28.5
h1 to h3	20.8	27.5	20.9	28.5
h2 to h3	20.9	28.6	22.1	29.9
h2 to h4	NA	NA	20.9	27.5
h3 to h4	NA	NA	21.6	28.5
h1 to h4	NA	NA	20.9	27.4

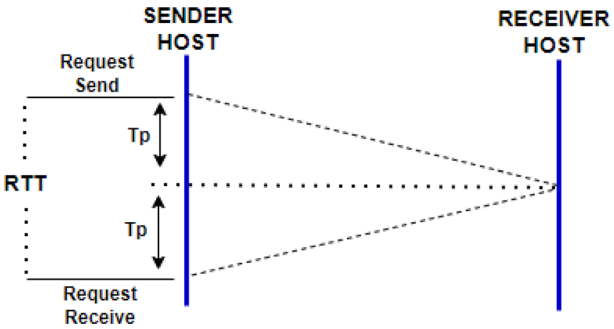


Fig. 13 RTT of the proposed topology

hops. Propagation delay is the length of time taken for a request to reach its destination. The propagation delay is usually the dominant component in RTT.

$$RTT = 2 * Propagation\,delay \tag{3}$$

As depicted in Fig. 14, the base RTTs of the default SDN topology [38] are estimated as 0.054 ms, 0.036 ms, 0.057 ms, 0.057 ms, 0.011 ms, and 0.047 ms. The normal RTTs of the proposed topology are estimated as 1.002 ms, 1 ms, 0.999 ms, 0.888 ms, 1.05 ms, and 0.918 ms. The long-run RTTs of the proposed topology are 9.45 ms, 9.345 ms, 7.336 ms, 9.668 ms, and 8.149 ms (Fig. 14).

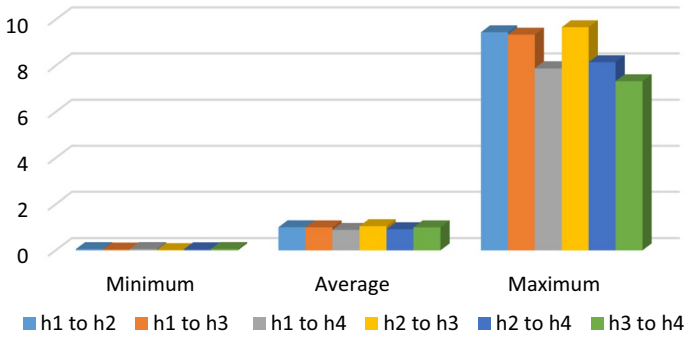


Fig. 14 RTT of the proposed topology

7.3.1 Comparison and Analysis of Proposed RTTs with Classical RTTs

In comparing the RTT between the proposed topology and the previous topology [38], several factors need to be considered. The proposed topology refers to the new network structure or configuration being suggested, while the previous topology represents the existing network design that is being replaced or improved upon.

In the case of the proposed topology, it has been designed with the optimization, it may aim to reduce network latency and minimize the number of network hops. By incorporating more direct routes or utilizing faster transmission mediums, the proposed topology has potentially decreased the overall round trip time as shown in Table 6. This improvement could lead to faster data transfer and reduced latency for end-users.

In contrast, the previous topology might have been developed without considering latency or with a different set of priorities. It could suffer from excessive network hops, suboptimal routing, or outdated infrastructure, all of which can contribute to increased round trip time. Consequently, users may experience slower data transmission and higher latency, resulting in a less efficient network.

Table 6 RTT comparison of the proposed and classical SDN topology

Nodes	Measures	RTT of classical SDN topology	RTT of proposed SDN topology
h1 to h2	Min	0.061	0.054
	Avg	1.411	1.002
	Max	10.305	9.45
h1 to h3	Min	0.055	0.036
	Avg	1.05	1
	Max	9.668	9.345
h2 to h3	Min	0.058	0.057
	Avg	0.918	0.888
	Max	8.149	7.889
h2 to h4	Min	NA	0.011
	Avg	NA	1.05
	Max	NA	9.668
h3 to h4	Min	NA	0.047
	Avg	NA	0.918
	Max	NA	8.149
h1 to h4	Min	NA	0.057
	Avg	NA	0.999
	Max	NA	7.336
h1 to h4	Min	NA	0.057
	Avg	NA	0.999
	Max	NA	7.336

8 Conclusion

SDN is cutting-edge computerization and also inculcated many changes in the industry of networking. SDN has separated both planes and made the controller a centralized hub. Therefore, the paper implemented the architecture of SDN using the RYU controller for the network traffic analysis. This work can use traffic routing in the environment of SDN to improve the network for better use from here on out. Subsequently, this work holds the analysis of traffic through performance assessment of the SDN controller to work on the network in different ways. Therefore, the proposed approach leverages the RYU controller, an open-source SDN controller framework, to collect and analyse network traffic data. By utilizing RYU's capabilities, we can dynamically monitor and capture network traffic statistics, such as bandwidth, throughput, packet counts, and RTT. These statistics provide valuable insights into network performance, and traffic patterns. This work also opens the door for many researchers working in the networking field. Being coded in Python, RYU is one of the most recommended Open-source SDN controllers. As RYU is scripted in the Python programming language, so only a few of the researchers are currently working on its performance evaluation. The consequence of the above-mentioned experimentations demonstrated that the RYU controller can likewise be viewed as one of the strong controllers as far as traffic engineering. The experimental work proved the viable outcomes in the data transmission starting with one node and then onto the next node in the proposed framework. In conclusion, this study highlights the significance of network traffic analysis in SDN environments and demonstrates the effectiveness of the Ryu controller for extracting valuable insights from network traffic data.

In future work, it would be recommended to consider several other topologies as well as other performance parameters to evaluate the RYU SDN controller performance. It may also focus on developing more sophisticated techniques to monitor and analyse network traffic in real time. This could involve implementing deep packet inspection (DPI) capabilities within the Ryu controller, allowing for detailed inspection and analysis of packet contents. Future research may explore integrating the RYU controller with security solutions, such as intrusion detection systems (IDS) and firewalls, to create a comprehensive security framework within the SDN environment. This integration can enable real-time traffic analysis for security purposes and allow for coordinated response mechanisms.

Author Contributions We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by both of us. We have all the required material used in the implementation and the available data/results are generated by the authors itself and not copied from somewhere else. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property. We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author.

Funding This study and authors received no fund.

Data availability There is no data available.

Declarations

Conflict of Interest We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

1. Bauerle Bass, S. (2003). How will Internet use affect the patient? A review of computer network and closed Internet-based system studies and the implications in understanding how the use of the Internet affects patient populations. *Journal of health psychology, 8*(1), 25–38.
2. Min, Y. H. (2003). A survey on computer network routing. *Chinese Journal of Computers, 26*(6), 641–649.
3. Mousa, A. H., Mohammed, N. T., & Mohammed, E. A. (2019). EFCNT: An evaluation framework for computer's network topologies. In *AIP Conference Proceedings*, October 24–25, 2019, Indonesia, *2144* (1), 50–64.
4. Ripeanu, M. (2001). Peer-to-peer architecture case study: Gnutella network. In Proceedings first international conference on peer-to-peer computing, August 27–29, 2001, Sweden, pp. 99–100. IEEE.
5. Santra, S., & Acharjya, P. P. (2013). A study and analysis on computer network topology for data communication. *International Journal of Emerging Technology and Advanced Engineering, 3*(1), 522–525.
6. Zhang, Y., Zhang, H. L., & Fang, B. X. (2004). A survey on Internet topology modeling. *Journal of Software, 15*(8), 1220–1226.
7. Lali, M., Mustafa, R., Ahsan, F., Nawaz, M., & Aslam, W. (2017). Performance evaluation of software defined networking vs. traditional networks. *The Nucleus, 54*(1), 16–22.
8. Bhardwaj, S., Panda, S. N., & Datta, P. (2020, March). Comparison and performance evaluation of software-defined networking controllers. In: *2020 international conference on emerging smart computing and informatics (ESCI)* (pp. 276–281). IEEE.
9. Zhu, M., Cao, J., Pang, D., He, Z., & Xu, M. (2015). SDN-based routing for efficient message propagation in VANET. *International Conference on Wireless Algorithms, Systems, and Applications, 23(2)* (pp. 788–797). Springer.
10. Hakiri, A., Gokhale, A., Berthou, P., Schmidt, D. C., & Gayraud, T. (2014). Software-defned networking: Challenges and research opportunities for future internet. *Computer Networks, 75*, 453–471.
11. Liatifis, A., Sarigiannidis, P., Argyriou, V., & Lagkas, T. (2023). Advancing sdn from openflow to p4: A survey. *ACM Computing Surveys, 55*(9), 1–37.
12. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems, 29*(7), 1645–1660.
13. Wang, X., Chen, X., Wang, Y., & Ge, L. (2020). An efficient scheme for SDN state consistency verification in cloud computing environment. *Concurrency and Computation: Practice and Experience, 32*(2), 5440–5449.
14. Masoudi, R., & Ghafari, A. (2016). Software defned networks: A survey. *Journal of Network and Computer Applications, 67*(4), 1–25.
15. Kreutz, D., Ramos, F. M. V., Verıssimo, P. E., Rothenberg, C. E., Azodolmolky, S., & Uhlig, S. (2015). Software-defned networking: A comprehensive survey. *Proceedings of the IEEE, 103*(1), 14–76.
16. Nunes, B. A. A., Mendonca, M., Nguyen, X. N., Obraczka, K., & Turletti, T. (2014). A survey of software-defned networking: Past, present, and future of programmable networks. *IEEE Communications Surveys & Tutorials, 16*(3), 1617–1634.
17. Jammal, M., Singh, T., Shami, A., Asal, R., & Li, Y. (2014). Software defned networking: State of the art and research challenges. *Computer Networks, 72*(1), 74–98.
18. Yang, M., Li, Y., Jin, D., Zeng, L., Wu, X., & Vasilakos, A. V. (2015). Software-defned and virtualized future mobile and wireless networks: A Survey. *Mobile Networks and Applications, 20*(1), 4–18.
19. Akyildiz, I. F., Lin, S. C., & Wang, P. (2015). Wireless software-defned networks (W-SDNs) and network function virtualization (NFV) for 5G cellular systems: An overview and qualitative evaluation. *Computer Networks, 93*(3), 66–79.
20. Alam, I., Sharif, K., Li, F., Latif, Z., Karim, M. M., Biswas, S., & Wang, Y. (2020). A survey of network virtualization techniques for internet of things using sdn and nfv. *ACM Computing Surveys (CSUR), 53*(2), 1–40.
21. Barakabitze, A. A., Ahmad, A., Mijumbi, R., & Hines, A. (2020). 5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges. *Computer Networks, 167*(2), 978–984.

22. Wang, Z., Ding, H., Li, B., Bao, L., Yang, Z., & Liu, Q. (2022). Energy efficient cluster based routing protocol for WSN using firefly algorithm and ant colony optimization. *Wireless Personal Communications, 125*(3), 2167–2200.
23. Wang, Z., Ding, H., Li, B., Bao, L., & Yang, Z. (2020). An energy efficient routing protocol based on improved artificial bee colony algorithm for wireless sensor networks. *IEEE Access, 8*, 133577–133596.
24. Chica, J. C. C., Imbachi, J. C., & Vega, J. F. B. (2020). Security in SDN: A comprehensive survey. *Journal of Network and Computer Applications, 159*(4), 102–118.
25. Caria, M., Jukan, A., & Hoffmann, M. (2013). A performance study of network migration to SDN-enabled traffic engineering. In *2013 IEEE Global Communications Conference (GLOBECOM)*, December 09–13, 2012, Atlanta, USA, pp. 1391–1396. IEEE.
26. Akyildiz, I. F., Lee, A., Wang, P., Luo, M., & Chou, W. (2016). Research challenges for traffic engineering in software defined networks. *IEEE Network, 30*(3), 52–58.
27. Wang, M. H., Chen, L. W., Chi, P. W., & Lei, C. L. (2017). SDUDP: A reliable UDP-Based transmission protocol over SDN. *IEEE Access, 5*, 5904–5916.
28. Yu, L., Wang, Q., Barrineau, G., Oakley, J., Brooks, R. R., & Wang, K. C. (2017, October). TARN: A SDN-based traffic analysis resistant network architecture. In *2017 12th International Conference on Malicious and Unwanted Software (MALWARE)*, October 11–14, 2017, USA, pp. 91–98. IEEE.
29. Bholebawa, I. Z., & Dalal, U. D. (2018). Performance analysis of SDN/OpenFlow controllers: POX versus foodlight. *Wireless Personal Communications, 98*(2), 1679–1699.
30. Guo, Y., Wang, Z., Yin, X., Shi, X., & Wu, J. (2017). Traffic engineering in hybrid SDN networks with multiple traffic matrices. *Computer Networks, 126*(1), 187–199.
31. Amin, R., Reisslein, M., & Shah, N. (2018). Hybrid SDN networks: A survey of existing approaches. *IEEE Communications Surveys & Tutorials, 20*(4), 3259–3306.
32. Singh, S., & Jha, R. K. (2019). SDOWN: A novel algorithm and comparative performance analysis of underlying infrastructure in software defined heterogeneous network. *Fiber and Integrated Optics, 38*(1), 43–75.
33. Mehr, S. Y., & Ramamurthy, B. (2019, December). An SVM based DDoS attack detection method for RYU SDN controller. In *Proceedings of the 15th international conference on emerging networking experiments and technologies*, December 19–20, Florida, US, pp. 72–73.
34. Bhatia, J., Dave, R., Bhayani, H., Tanwar, S., & Nayyar, A. (2020). Sdn-based real-time urban traffic analysis in vanet environment. *Computer Communications, 149*, 162–175.
35. Priya, A. V., & Radhika, N. (2019). Performance comparison of SDN OpenFlow controllers. *International Journal of Computer Aided Engineering and Technology, 11*(4–5), 467–479.
36. Badotra, S., & Panda, S. N. (2019). Evaluation and comparison of OpenDayLight and open networking operating system in software-defined networking. *Cluster Computing, 3*(1), 1–11.
37. Queiroz, W., Capretz, M. A., & Dantas, M. (2019). An approach for SDN traffic monitoring based on big data techniques. *Journal of Network and Computer Applications, 131*(1), 28–39.
38. Islam, M. T., Islam, N., & Al Refat, M. (2020). Node to Node Performance Evaluation through RYU SDN Controller. *Wireless Personal Communications*. https://doi.org/10.1007/s11277-020-07060-4
39. Kurungadan, B., & Abdrabou, A. (2022). Using software-defined networking for data traffic control in smart cities with WiFi coverage. *Symmetry, 14*(10), 2053.
40. Bhardwaj, S., & Panda, S. N. (2022). Performance evaluation using ryu sdn controller in software-defined networking environment. *Wireless Personal Communications, 122*(1), 701–723.
41. Raja, N. M., & Vegad, S. (2023). An empirical study for the traffic flow rate prediction-based anomaly detection in software-defined networking: A challenging overview. *Social Network Analysis and Mining, 13*, 72.
42. Shukla, P. K., Maheshwary, P., Subramanian, E. K., Shilpa, V. J., & Varma, P. R. K. (2023). Traffic flow monitoring in software-defined network using modified recursive learning. *Physical Communication, 57*, 101997.
43. Bholebawa, I. Z., & Dalal, U. D. (2016). Design and performance analysis of OpenFlow-enabled network topologies using mininet. *International Journal of Computer and Communication Engineering, 5*(6), 419.
44. Tootoonchian, A., Gorbunov, S., Ganjali, Y., Casado, M., & Sherwood, R. (2012). On controller performance in software-defned networks. In *Presented as part of the 2nd USENIX workshop on hot topics in management of internet, cloud, and enterprise networks and services*, March 29–30, 2011, Boston, pp. 22–30.
45. Al-Rubaie, N. R. O., Kamel, R. N. N., & Alshemari, R. M. (2023). Simulating fog computing in OMNeT++. *Bulletin of Electrical Engineering and Informatics, 12*(2), 979–986.

46. Peyman, M., Fluechter, T., Panadero, J., Serrat, C., Xhafa, F., & Juan, A. A. (2023). Optimization of vehicular networks in smart cities: From agile optimization to learnheuristics and simheuristics. *Sensors, 23*(1), 499.
47. Leon, J., Aydeger, A., Mercan, S., & Akkaya, K. (2023). SDN-enabled vehicular networks: Theory and practice within platooning applications. *Vehicular Communications, 39*, 100545.
48. Singh, P. K., Sharma, S., Nandi, S. K., & Nandi, S. (2019). Multipath TCP for V2I communication in SDN controlled small cell deployment of smart city. *Vehicular communications, 15*, 1–15.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Shanu Bhardwaj is currently pursuing PhD in Department of Computer Science & Engineering at Delhi Technological University, Delhi, India. She has completed her B. Tech in Computer Science and Engineering at Kurukshetra University, Kurukshetra (India) and M.E. at Chitkara University, Punjab (India). Her research areas are Software-Defined Networking (SDN), Internet of Things (IoT), Wireless sensor network (WSN) and Network security. She has completed her internship on "Application of Big Data in Construction Law" from Tamkang University, Taiwan.

Ashish Girdhar is currently working as Assistant Professor in Department of Computer Science & Engineering at Delhi Technological University since December 2020. He has total of more than 9 years of Teaching Experience. He has done his PhD in area of Image Processing from Thapar Institute of Engineering and Technology, Patiala. He has 7 SCI/SCIE Publications and few other publications. His major area of research are Image Processing and Machine Learning.

Nonlinear Regression Analysis of Rut Profile Data for Optimal Data Storage and Efficient Terrain Condition Analysis

Manoj Kumar Kalra, Ashutosh Trivedi, and Sanjay Kumar Shukla

***Abstract*— During various operations of military and civil importance, the vehicles need to move in different off-road conditions. Several soil parameters that differ in spatial and temporal domains impact the possibility of vehicular movement. Conventional evaluation of the prevalent state of soil being cumbersome and time-consuming, the information extracted from the rut profile of the vehicle is used. Several techniques ranging from manual to advanced laser scanners are there to monitor this profile. The laser scan contains a vast point cloud dataset. It brings in the data storage and retrieval issue for an efficient computational framework for real-time mobility guidance. Little research exists for addressing the storage issue of this sort of rut profile data. Here, rut profile data captured manually in different field running conditions are used for studying the optimization aspects. Various mathematical formulations are analyzed, and nonlinear regression analysis is attempted on the commonly observed rut shapes. Using the proposed modified bell-shaped function on these common shapes, the $R2$ value moved above 0.98. It could fetch an additional compression of over 80% on straight patches and 71% on turnings. Although complex scenarios could be dealt with separately, an optimal representation of the most common rut shapes can aid the development of an efficient decision support system.**

***Index Terms*— Off-road conditions, optimization, regression analysis, rut profile, soil condition, vehicular movement.**

I. Introduction

THE poor performance of military vehicles during World War II led to the beginning of research in trafficability at the Waterways Experiment Station (WES) in 1945 by U.S. Army Engineer [1]. The characterization of vehicular movement on different unpaved terrain features is a challenging task. The possibility of vehicular movement depends upon the surface topographical features and the condition of the underlying soil. Wong [2] and Saarilahti [3] and others considered the effect of these obstacles both empirically and mathematically.

The soil is a dynamic material by nature, and parameters like density, moisture, constituents, etc. influence its state. Evaluating these parameters and their effect by conventional geotechnical instruments is laborious and time-consuming. American Society of Agricultural Engineers (ASAE) standards [4] give the procedure for estimating the strength of the soil using a cone penetrometer. However, this tool is suitable for a limited area only and is not convenient to map large areas. Various authors made attempts to explore alternate means for evaluating the soil condition. One such way is to observe the rut profile formed by the movement of vehicles on the unpaved ground. As strength decreases, the rut depth increases. Vehicle immobilization condition occurs when the rut depth is larger than the vehicle clearance height [5].

However, the rut profile is considered a better measure for mapping the influence of terrain on turnings and conditions like sandy terrain. A study is made on the changes in rut volume by vehicular movements on turnings [6]. The equipment for mapping rut profiles varies from manual profilometers [5] to several advanced profilers [7]. Different technologies in use for rut measurement include ultrasonic sensors, point lasers, scanning lasers, and optical sensors. The manual profilometer is employed to collect data for point locations; however, it is not suitable for continuous measurements of changes in terrain properties.

Photogrammetric methods, unmanned aerial vehicle (UAV), and drone platforms are employed to map the rut profiles of the soil [8], [9], [10], [11]. A forest machine-based light detection and ranging (LiDAR) sensor is used by [12] to demonstrate it as an efficient and reliable way of collecting the rut depth data. The use of mobile laser profilometry for detecting forest road damage and its quantification is presented [13]. The profiler used in this study had eight scanner units, which could cover the whole width of the road with 1000–5000 profiles per second.

Thus, it is evident how much point cloud data processing is required to create a profile and do further analysis. The importance of rut measurement for applications like controller design and implementation is highlighted by [14]. Applications of LiDAR have been advancing to mobile-based platforms. An iPhone-12 is used to present LiDAR-derived snow depth estimation [15]. This work which used a lightweight LiDAR system brings out the need for research on reducing the computational load for efficient decisions.

The data collected by such scanners is vast. It needs heavy resources for storage, retrieval, and processing. Limited attempts are there to make use of the rut shape mathematically. Limited attempts are there to make use of the rut shape

Manuscript received 30 April 2023; revised 26 June 2023; accepted 2 July 2023. Date of publication 14 August 2023; date of current version 29 August 2023. *(Corresponding author: Manoj Kumar Kalra.)*

Manoj Kumar Kalra is with the Defence Geoinformatics Research Establishment (DGRE), DRDO, Chandigarh 160036, India, and also with the Department of Civil Engineering, Delhi Technological University, Delhi 110042, India (e-mail: kalramk@rediffmail.com).

Ashutosh Trivedi is with the Department of Civil Engineering, Delhi Technological University, Delhi 110042, India.

Sanjay Kumar Shukla is with the School of Engineering, Edith Cowan University, Perth, WA 6027, Australia.

Digital Object Identifier 10.1109/LGRS.2023.3303134

1558-0571 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:29:07 UTC from IEEE Xplore. Restrictions apply.

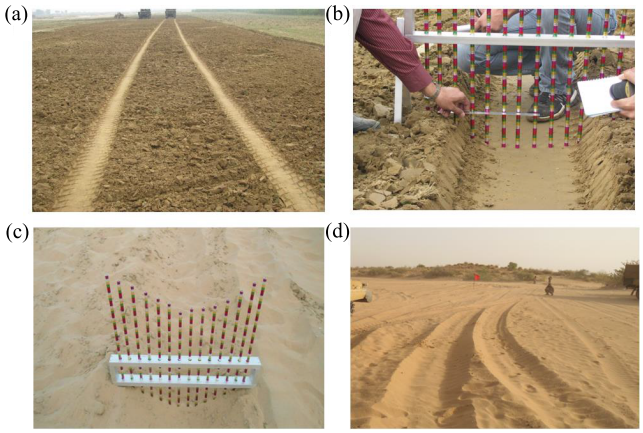


Fig. 1. Different shapes of rut profiles as observed in field. (a) Shallow rut in alluvial soils of medium consistency. (b) Deep rectangular rut on soils of low consistency. (c) Conical rut profile on frictionless soils of deserts. (d) Asymmetric rut shape created by vehicles on turnings.

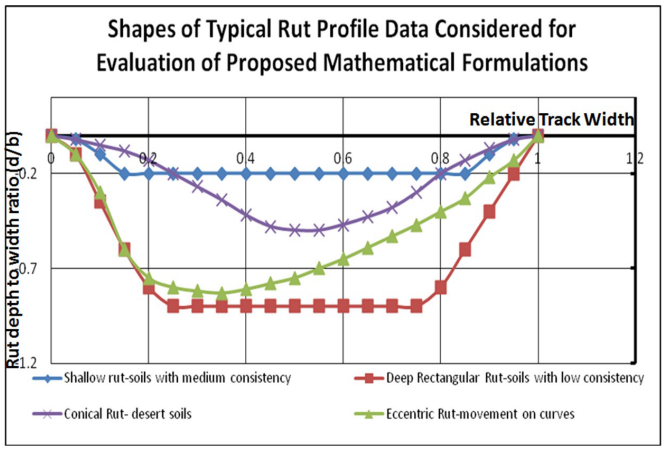


Fig. 2. Shapes of typical rut profile data considered for evaluation of proposed mathematical formulations.

mathematically. In the study on the rut following robot, [16] modeled the previously formed rut shape using a second-order polynomial.

Defining the shape of the deformed geometry is attempted in other applications too. A general equation for the deformed geometry of geosynthetic is given [17] that combines the two shapes to compute strain. A regression-based solution for estimating the settlement of reinforced soil foundation is attempted [18].

This letter presents a comparative analysis of some suitably selected mathematical formulations that could fit the typical shapes of ruts profiles. Nonlinear regression analysis is attempted on the observed data to identify the most suitable equation representing the rut profile in different conditions. The letter discusses the suitability of these mathematical formulations for depicting rut profiles in different scenarios to achieve optimal data storage.

II. Field Observation of Rut Profiles

The shape of the rut profile is typical of different terrain features, soil conditions, and vehicle running conditions. A manual rut profiler is used for this study to collect rut profile data in different kinds of terrain scenarios. This profiler consists of rods placed at intervals of 5 cm. The color marking of width 1 cm each is made on each rod to facilitate the measurement of rut depth. The authors investigated rut profiles created in different unpaved terrain conditions of deserts and alluvial terrain in both dry and wet seasons. These data are used for further analysis as given in this letter.

As illustrated in Fig. 1, the shallow rut profile gets formed on alluvial soils of medium consistency. However, a deeper rut gets formed as the consistency of soil decreases.

The rut profile in loose sandy dunal soils of deserts takes a conical shape as the sand from the edges pours in as the rut profile. Further, when the vehicle turns on curves, the rut shape becomes asymmetric due to lateral loads.

There are other possible shapes too, but the following typical cases of rut shapes being most common during the movement in different off-road scenarios are considered here.

TABLE I

Depth-to-Width Ratio of Rut Under Different Soil and Movement Conditions

Relative track width	Rut profiles by vehicles in different soil and movement conditions			
	On straight Paths			On Turnings by vehicles
	Alluvial soil-medium consistency	Alluvial soil-low consistency	Desert soil -sand dunes	
b/B	d/B	d/B	d/B	d/B
x	y1	y2	y3	y4
0	0	0	0	0
0.05	-0.02	-0.1	-0.1	-0.02
0.1	-0.1	-0.35	-0.3	-0.05
0.15	-0.2	-0.6	-0.6	-0.08
0.2	-0.2	-0.8	-0.75	-0.13
0.25	-0.2	-0.9	-0.8	-0.2
0.3	-0.2	-0.9	-0.82	-0.27
0.35	-0.2	-0.9	-0.83	-0.34
0.4	-0.2	-0.9	-0.81	-0.42
0.45	-0.2	-0.9	-0.78	-0.48
0.5	-0.2	-0.9	-0.75	-0.5
0.55	-0.2	-0.9	-0.7	-0.5
0.6	-0.2	-0.9	-0.65	-0.47
0.65	-0.2	-0.9	-0.59	-0.43
0.7	-0.2	-0.9	-0.53	-0.38
0.75	-0.2	-0.9	-0.47	-0.3
0.8	-0.2	-0.8	-0.4	-0.2
0.85	-0.2	-0.6	-0.33	-0.13
0.9	-0.1	-0.4	-0.22	-0.07
0.95	-0.02	-0.2	-0.13	-0.02
1	0	0	0	0

Case 1: Shallow rectangular shape rut in soils of medium consistency.

Case 2: Deep rectangular/ trapezoidal shape rut in alluvial soils of low consistency.

Case 3: Conical shape rut on the desert soils.

Case 4: Asymmetric rut profile in soils on turnings by vehicle.

The typical cases depicted in Fig. 2 are considered in this study to explore the optimal representation of these profiles using some mathematical formulations.

The authors normalized the cross-sectional rut profile to account for the variation of the rut profile at different places. Thus the depth 'y' representing 'd/B' at different sections 'x' across the profile representing 'B' is used. Here 'd' is the rut depth at different locations 'b' across the profile, and 'B' represents the total rut width. The consolidated rut profile data taken for study for all four cases are given in Table I.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:29:07 UTC from IEEE Xplore. Restrictions apply.

III. Development of Mathematical Formulations for Rut Profile

The authors explored many possible mathematical equations to develop a suitable mathematical formulation that explains rut profiles in varied scenarios. Based on the possibility of generating the most common types of shapes by changing the parameters, the authors chose the mathematical formulations. The equations thus chosen include polynomial equations of varying degrees, a generalized Gaussian shape equation, Fourier transformation, and a modified bell-shaped equation.

As a common approach, the polynomial fit of different degrees is resorted for defining the rut shape. The polynomial equation thus chosen is as follows:

$$y = f(x) = \sum_{i=0}^{n} a_i x^i. \tag{1}$$

The degree of the polynomial equation improves the curve fitting for the most common shapes. As such, regression fit used polynomial equations starting with a quadratic equation and going up to the fourth degree to define the rut shapes.

The rut shape in frictionless soils resembles an inverted shape of the Gaussian distribution of probability function. The generalized form of the equation for the Gaussian shape curve is therefore included and given as follows:

$$f(x) = a_1 . e^{-\left|\left(\frac{x-a_2}{a_3}\right)\right|^{a_4}}. \tag{2}$$

Another way of describing the rut shape is similar to converting any harmonic function using Fourier transformation. This method approximates the curve using a combination of various sine and cosine curves of different frequencies, as given below:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{nx}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{nx}{L}\right). \tag{3}$$

Here, the coefficients are

$$a_0 = \frac{1}{L} . \int_{-L}^{L} f(x) dx \tag{4}$$

$$a_n = \frac{1}{L} . \int_{-L}^{L} f(x) \cos\left(\frac{nx}{L}\right) dx \tag{5}$$

$$b_n = \frac{1}{L} . \int_{-L}^{L} f(x) \sin\left(\frac{nx}{L}\right) dx. \tag{6}$$

The function combines various frequencies and amplitudes of the sine and cosine curves. In the present case, different shapes of rut are represented by considering up to three terms (n) of the above equation.

The rut shapes on soils of different consistency also resemble the shape of a bell function, and accordingly, the bell-shaped function is thus chosen and given as follows:

$$f(x) = \frac{a_1}{1 + \left|\left(\frac{x-a_2}{a_3}\right)\right|^{a_4}}. \tag{7}$$

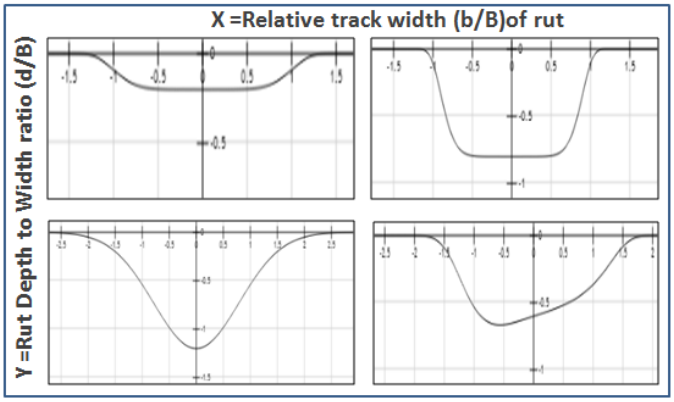


Fig. 3. Different shapes as generated using proposed mathematical formulations by varying the input parameters to match the commonly observed rut shapes in various terrain types.

Here, the coefficient a_1 impacts the variation in rut width and depth. The coefficient a_2 determines the center, a_3 represents the rut width, and a_4 the side slope of the membership function curve. Therefore, the use of a modified form of the generalized bell equation, which combines variation in the side slopes, is considered and is given below:

$$f(x) = \frac{a_1}{1 + \left|\left(\frac{x-a_2}{a_3}\right)\right|^{a_4} + \left|\left(\frac{x-a_2}{a_5}\right)\right|^{a_6}}. \tag{8}$$

All these equations use the normalized rut shapes. They represent the dimensionless ratio of rut depth to its width at various relative locations across the rut profiles. The parameters $a_0, a_1, a_2, a_3, a_4, a_5$, and a_6 are the coefficients of the proposed equations.

IV. Results and Discussion

The suitability of proposed equations is first analyzed based on regression fit on rut data observed in different scenarios. As a general observation, by taking typical values of the variables, one can generate different shapes using any of the proposed mathematical formulations. Fig. 3 illustrates the shapes of curves as obtained using mathematical formulations.

It indicates that using any of the proposed mathematical formulations, one can generate the shape of rut profiles in scenarios observed in the field. It also leads to the inference that the selected equations are reasonable and logical. These equations can thus be used as an alternate to represent the rut profiles data. The potential use could be to achieve the optimal data storage requirements as it becomes mandatory in systems using point cloud as given by [13].

Further, to evaluate the effectiveness of different proposed mathematical formulations, authors employed non-linear regression analysis. The online desmos graphing tool facilitated the computation of regression coefficients and the degree of fit. Table II displays the result of regression analysis representing the efficacy of different mathematical formulations in various conditions.

"Here, the compression ratio (C) represents the storage efficiency. It is the ratio of the number of original datasets to the number of parameters needed to represent the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:29:07 UTC from IEEE Xplore. Restrictions apply.

TABLE II
REGRESSION FIT RESULTS FOR RUT PROFILE IN DIFFERENT TEST CASES

S No.	Type of Curve / Regression Coefficients	Regression fit result for various types of curves a_0	a_1	a_2	a_3	a_4	a_5	a_6	R^2	C
Case1: Rut profile in alluvial soils of medium consistency										
1	Polynomial curve of 2nd degree	-0.022	-0.834	0.8338	-	-	-	-	0.806	7.0
2	Polynomial curve of 3rd degree	0.022	-0.834	0.8338	1E-15	-	-	-	0.806	5.25
3	Polynomial curve of 4th degree	0.026	-2.074	6.7335	-9.32	4.66	-	-	0.924	4.2
4	Generalised Gaussian shape equation	-	-0.202	0.5	-0.414	14.23	-	-	0.991	5.25
5	Fourier Series equation	-0.006	-0.225	0	-0.064	0	0.021	0	0.941	3.0
6	Generalised Bell shape equation	-	-0.201	0.5	0.401	22.58			0.998	5.25
7	Modified Generalised bell shaped Equation	-	-0.201	0.5	5	30	-0.401	22.58	0.998	3.5
Case2: Rut profile in alluvial soils of low consistency										
1	Polynomial curve of 2nd degree	-0.027	-3.894	3.8599	-	-	-	-	0.934	7.0
2	Polynomial curve of 3rd degree	-0.02	-3.991	4.1095	-0.166	-	-	-	0.934	5.25
3	Polynomial curve of 4th degree	0.095	-6.933	18.109	-22.28	11.06	-	-	0.970	4.2
4	Generalised Gaussian shape equation	-	-0.907	0.5046	-0.407	6.495	-	-	0.996	5.25
5	Fourier Series equation	-0.868	0.4301	0.0156	-0.447	-3E-04	0.879	0.018	0.994	3.0
6	Generalised Bell shape equation	-	-0.9	0.5039	0.38	9.645	-	-	0.993	5.25
7	Modified Generalised bell shaped Equation	-	-0.906	-2.806	-3.692	85.57	2.928	-79.45	0.996	3.5
Case 3: Rut profile on sand dunes of desert soil										
1	Polynomial curve of 2nd degree	0.118	-2.118	2.0596	-	-	-	-	0.887	7.0
2	Polynomial curve of 3rd degree	0.072	-1.488	0.4453	1.076				0.903	5.25
3	Polynomial curve of 4th degree	-0.028	1.0727	-11.74	20.33	-9.625			0.993	4.2
4	Generalised Gaussian shape equation	-	-0.497	0.5286	-0.285	2.369			0.997	5.25
5	Fourier Series equation	-0.119	-0.219	0.0108	0.042	0.023	0.119	-0.026	0.999	3.0
6	Generalised Bell shape equation	-	-0.48	0.53	0.243	3.82	-	-	0.990	5.25
7	Modified Generalised bell shaped Equation	-	-0.572	1.2327	0.476	-5.815	-0.924	11.45	0.998	3.5
Case 4: Rut profile on turnings in soils										
1	Polynomial curve of 2nd degree	-0.117	-2.903	3.1181	-	-	-	-	0.887	7.0
2	Polynomial curve of 3rd degree	0.043	-5.087	8.7144	-3.731	-	-	-	0.970	5.25
3	Polynomial curve of 4th degree	0.089	-6.286	14.419	-12.74	4.506	-	-	0.978	4.2
4	Generalised Gaussian shape equation	-	-0.768	0.4547	-0.381	3.884	-	-	0.912	5.25
5	Fourier Series equation	-0.334	-0.244	-0.167	-0.192	0.023	0.345	0.007	0.989	3.0
6	Generalised Bell shape equation	-	-0.758	0.4503	0.341	5.755	-	-	0.911	5.25
7	Modified Generalised Bell shaped Equation	-	1.1156	1.0156	0.324	-1.368	-0.881	24.06	0.996	3.5

profile using mathematical formulation. The value of (C) in Table II indicates that one can achieve significant levels of data compression over conventional techniques using the proposed mathematical models. In the currently used manual observation of profiles, the compression ratio for the straight patches is 5.25. The %age of data compression calculated as (1-1/C) $\times$ 100 indicates that one can achieve more than 80% compression on the straight patches. The value computes for areas on turning is 71%, while the accuracy level is maintained

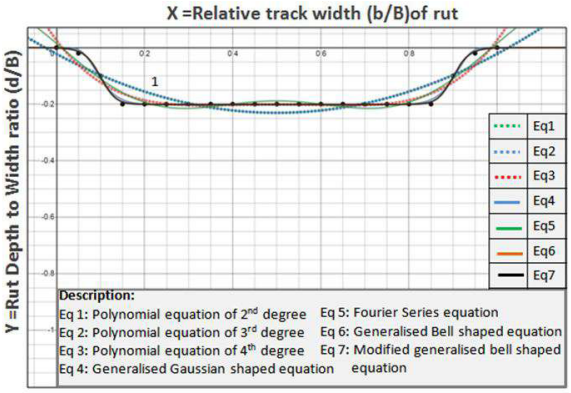


Fig. 4. Typical regression fit curves on the rut profile data of medium consistency soil.

better than 99% in both cases. This additional data compression could pave the way for efficient data storage and set the basis for an onboard decision aid. The computational aspects will depend upon the selected mathematical formulation, regression type, algorithm design, parallel processing, multithreading aspects, etc. Here, the purpose is to explore the potential of mathematical formulations in defining rut shapes to address the issues related to real-time applications."

Fig. 4 shows the typical regression fit curves on the rut profile data of medium-consistency soils. The figure shows the relative comparison of various mathematical formulations on the observed data.

The graph containing the data and different regression curves indicate that the curves using the proposed mathematical formulations follow the observed data for almost all the test cases. However, for identifying the most suitable predictive equation of the rut profile, the comparative analysis reveals further detail.

The analysis of the regression fit shown in the table also indicates that in almost all the cases, the regression fit improves with the increased degree of the polynomial. One can plan to improvise the utility of rut profiles by representing better the rut shape as given by [16].

The study brings out an analysis of an equivalent number of variables used in the regression curves. The comparison given for the third-degree polynomial fit with other mathematical formulations reveals valuable inferences. A considerable improvement is there in the generalized Gaussian shape function than the polynomial equations when movement is on the straight patches.

The goodness of fit is relatively less along the curves. The higher degree equations are, therefore, employed for a better fit. The analysis of regression fit results also indicates that while using the proposed modified generalized bell-shaped equation, the results are much better in almost all the scenarios. However, the function needs six input parameters to represent the rut shape.

From the above analysis, some useful inferences are drawn. The generalized Gaussian form or bell-shaped function may be advantageous for obtaining precision and optimal storage on straight patches. However, the modified generalized bell-shaped equation is more accurate on the curves.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:29:07 UTC from IEEE Xplore. Restrictions apply.

V. Conclusion and Future Work

Results suggest that one can employ mathematical formulations to represent the rut shapes for optimal data storage. As rut profile is characteristic of soil condition, its mathematical representation can help to devise rules for interpreting soil and associated trafficability conditions. The associated interpretation based on the soil condition can be more efficient. One can employ a generalized Gaussian or bell-shaped equation to represent the rut profile on the straight patches. However, on turnings, the proposed modified bell-shaped equation leads to better precision and optimization.

An interesting analysis would be to explore some complex cases in the future. For instance, one may encounter a class wherein two separate rut profiles are blended to create a single rut class. Here some additional mathematical representations may make the profile representation more generic and can address some more rut classes.

In situations where complexity in rut morphologies is there, such as on crossings and different pathways, efforts need to be there to represent the rut profile in a better way. The present work focuses on the new ways to represent optimally the rut profile in different terrain conditions using mathematical equations. It will be a step toward developing an efficient onboard system to store and infer different rut profile-based properties. The categorization of rut profiles, mapping and characterizing the spatial variation of soil type and its condition, and inference about the trafficable state of terrain in near real-time mode are some directly related benefits. In future activities, one can explore the machine learning approach to train the system with rut profile data of varied field scenarios. It can make predictive analysis about various terrain properties and classes in different field running conditions.

The standard cases given in this study are observed prominently in actual off-road conditions. The proposed method indicates the utility of representing the rut shapes mathematically to achieve optimization. This study is a step toward bringing efficiency in addressing various issues of vehicular mobility during operational needs.

References

[1] A. A. Rula and C. J. Nuttall, "An analysis of ground mobility models (ANAMOB)," U.S. Army Corps Eng. Waterways Exp. Station, Vicksburg, MS, USA, Tech. Rep. M-71-4, 1971.

[2] J. Y. Wong, *Theory of Ground Vehicles*. New York, NY, USA: Wiley, 1978.

[3] M. Saarilahti, "Soil interaction model-development of a protocol for ecoefficient wood harvesting on sensitive sites (ECOWOOD),"Dept. Forest Resour. Manag., Forest Soil Properties, Appendix 8, Univ. Helsinki, Helsinki, Finland, Tech. Rep., 2002.

[4] S. Joseph, "Procedures for using and reporting data obtained with the soil cone penetrometer," ASAE, Washington, DC, USA, Tech. Rep. ASAE EP542, 1999.

[5] R. T. Affleck, "Disturbance measurements from off-road vehicles on seasonal terrain," Engineer Res. Develop. Center, U.S. Army Corps Eng., Tech. Rep. ERDC/CRREL TR-05e12, 2005, Accessed: May 11, 2021. [Online]. Available: http://hdl.handle.net/11681/5315

[6] K. Liu, "Influence of turning on military vehicle induced rut formation," Ph.D. dissertation, Graduate School Univ. Tennessee, Knoxville, TN, USA, 2009. [Online]. Available: http://trace.tennessee.edu/utkgraddiss/620

[7] R. Mallela and H. Wang, "Harmonising automated rut depth measurements—Stage 2," Land Transp. New Zealand Res., New Zealand, OC, Australia, Tech. Rep. 277, 2006.

[8] P. Nevalainen et al., "Estimating the rut depth by UAV photogrammetry," *Remote Sens.*, vol. 9, no. 12, p. 1279, Dec. 2017.

[9] E. Marra, M. Cambi, R. Fernandez-Lacruz, F. Giannetti, E. Marchi, and T. Nordfjell, "Photogrammetric estimation of wheel rut dimensions and soil compaction after increasing numbers of forwarder passes," *Scandin. J. Forest Res.*, vol. 33, no. 6, pp. 613–620, Aug. 2018.

[10] E. Marra, R. Wictorsson, J. Bohlin, E. Marchi, and T. Nordfjell, "Remote measuring of the depth of wheel ruts in forest terrain using a drone," *Int. J. Forest Eng.*, vol. 32, no. 3, pp. 224–234, Sep. 2021.

[11] E. Marra et al., "Impacts of wood extraction on soil: Assessing rutting and soil compaction caused by skidding and forwarding by means of traditional and innovative methods," *Eur. J. Forest Res.*, vol. 141, no. 1, pp. 71–86, Feb. 2022.

[12] A. Salmivaara et al., "Wheel rut measurements by forest machine-mounted LiDAR sensors–accuracy and potential for operational applications?" *Int. J. Forest Eng.*, vol. 29, no. 1, pp. 41–52, Jan. 2018.

[13] M. Ferenčík, M. Kardoš, M. Allman, and Z. Slatkovská, "Detection of forest road damage using mobile laser profilometry," *Comput. Electron. Agricult.*, vol. 166, Nov. 2019, Art. no. 105010.

[14] R. He, C. Sandu, A. K. Khan, A. G. Guthrie, P. S. Els, and H. A. Hamersma, "Review of terramechanics models and their applicability to real-time applications," *J. Terramechanics*, vol. 81, pp. 3–22, Feb. 2019.

[15] F. King, R. Kelly, and C. G. Fletcher, "Evaluation of LiDAR-derived snow depth estimates from the iPhone 12 pro," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[16] C. Ordonez, O. Y. Chuy, E. G. Collins, and X. Liu, "Rut tracking and steering control for autonomous rut following," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2009, pp. 278–2775.

[17] S. K. Shukla and N. Sivakugan, "A general expression for geosynthetic strain due to deflection," *Geosynthetics Int.*, vol. 16, no. 5, pp. 402–407, Oct. 2009.

[18] M. Khosrojerdi, M. Xiao, T. Qiu, and J. Nicks, "Nonlinear equation for predicting the settlement of reinforced soil foundations," *J. Geotechnical Geoenvironmental Eng.*, vol. 145, no. 5, p. 2027, May 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:29:07 UTC from IEEE Xplore. Restrictions apply.

ACS OMEGA

Open Access

This article is licensed under CC-BY-NC-ND 4.0

http://pubs.acs.org/journal/acsodf

Article

Novel Enzymatic Biosensor Utilizing a MoS_2/MoO_3 Nanohybrid for the Electrochemical Detection of Xanthine in Fish Meat

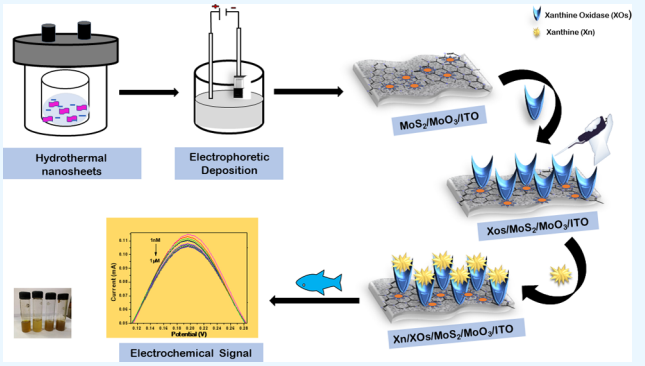
Prateek Sharma,* Deeksha Thakur, and Devendra Kumar

Cite This: *ACS Omega* 2023, 8, 31962−31971

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: A rapid, reliable, and user-friendly electrochemical sensor was developed for the detection of xanthine (Xn), an important biomarker of food quality. The developed sensor is based on a nanocomposite comprised of molybdenum disulfide-molybdenum trioxide (MoS_2/MoO_3) and synthesized using a single-pot hydrothermal method. Structural analysis of the MoS_2/MoO_3 nanocomposite was conducted using X-ray diffraction (XRD) and Raman spectroscopy, while its compositional properties were evaluated through X-ray photoelectron spectroscopy (XPS). Morphological features were observed using scanning electron microscopy (SEM) and transmission electron microscopy (TEM). Two-dimensional (2D) MoS_2 offers advantages such as a high surface-to-volume ratio, biocompatibility, and strong light−matter interaction, whereas MoO_3 serves as an effective electron transfer mediator and exhibits excellent stability in aqueous environments. The enzymatic biosensor derived from this nanocomposite demonstrates remarkable cyclic stability and a low limit of detection of 64 nM. It enables rapid, reproducible, specific, and reproducible detection over 10 cycles while maintaining a shelf life of more than 5 weeks. These findings highlight the potential of our proposed approach for the development of early detection devices for Xn.



Downloaded via ESS INFLIBNET PCA 4 on September 5, 2023 at 09:07:38 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Keeping fish meat fresh is a significant challenge for the food and pharmaceutical industries. Because fish meat is perishable, it requires careful preservation to resolve food safety concerns and preserve product excellence. To preserve the safety and quality of seafood products, keeping the freshness of fish meat is crucial in the food sector.[1] Once fish is handled or stored improperly, dangerous bacteria like Listeria, Salmonella, and Vibrio can quickly proliferate and pose a risk for foodborne illnesses for consumers. The industry uses a variety of strategies to address this problem, including proper storage, temperature management, and the application of preservatives. The pharmaceutical industry uses fish meat as a valuable ingredient in the manufacturing of pharmaceuticals like fish oil supplements and medications made from fish. In order to maintain the effectiveness and safety of these products, the freshness of the fish meat is essential. Thus, ensuring the freshness of fish meat stands as a substantial challenge in both the food and pharmaceutical sectors, necessitating adherence to proper handling, storage, and processing methods to guarantee the safety and quality of fish-based merchandise.

When a fish is sacrificed, the presence of endogenous enzymes in the intestine and the muscle tissues causes the decomposition of nucleotides and other autocatalytic reactions start immediately.[2,3] The rapid decomposition of adenosine-5-triphosphate (ATP) into adeonosine-5-di-phosphate (ADP) and further breaks into adenosine-5-monophosphate (AMP), inosine-5-monophosphate (IMP), inosine (I), hypoxanthine (HXn), and xanthine (Xn), respectively.[2,4,5] At the time of death, the autolysis causes the rotting of fish by ATP degradation to Xn (3,7-dihydro-purine-2,6-dione), which later ends in the formation of uric acid (UA) by the oxidation of Xn.[6,7] The compound IMP majorly contributes toward the taste of the meat, and its degradation to HXn and Xn leads to the bitter taste of the fish meat.[8] The level of chief metabolite Xn in the ATP degradation process in the dead fish increases with the storage time of fish meat.[9] Therefore, keeping track of the Xn level in the fish meat can be used as an important parameter to determine the freshness of the fish meat.[8]

Xn is a purine base that is found in various biological samples, including urine, blood, and cerebrospinal fluid. The Xn finally breaks into the UA, which ultimately breaks into the blood and passes through the kidney to come out as urine. Elevated levels of Xn can indicate certain medical conditions,

Received: May 29, 2023
Accepted: August 8, 2023
Published: August 22, 2023

© 2023 The Authors. Published by American Chemical Society

such as renal failure, diabetes influencer, hyperuricemia, liver, gout disease, cerebral ischemia, xanthinuria, and cardiac diseases.[10−14] Therefore, sensitive and accurate methods for xanthine sensing are important for disease diagnosis and monitoring. There are several methods for xanthine sensing, including enzymatic and nonenzymatic approaches. Enzymatic methods involve the use of xanthine oxidase (XOD), an enzyme that catalyzes the conversion of xanthine to uric acid. Nonenzymatic methods involve the use of various materials that can selectively bind to xanthine and generate a measurable signal. Overall, xanthine sensing is an important area of research with many potential applications in medical diagnostics and monitoring. Further research is needed to optimize the performance of xanthine sensors and to develop practical applications for disease diagnosis and monitoring.

Previously, many techniques have been deployed to detect Xn and its metabolites, like spectrophotometry, mass spectrometry, chemo/electroluminescence, high-performance liquid chromatography, and capillary electrophoresis biosensors.[15−21] The traditional techniques were supposedly time-consuming and had a limited range of detection toward the targeted analyte, which made researchers look out for other options. One such popular alternative is the electrochemical biosensors, which have an exquisite range of detection, simple processing, and high sensitivity and selectivity.[22−28] Till now, many attempts have been made to immobilize the enzymes onto the electrode's surface by means of adsorption, covalent, and intermediate molecular linkage, which affects the sensing ability of the electrochemical sensors.[29,30] In the last decade, nanoparticles have been extensively used due to their unique properties and great potential in the field of biosensing.[31−37] One such class of popular nanomaterials is transition metal dichalcogenides (TMDs). These are the two-dimensional (2D) nanomaterials having a high surface-to-volume ratio and good electronic response, which could be employed as potent electrodes for Xn detection. On the other side, the layered metal oxides have also gained attention as a new class of van der Waals material.

The combination of the materials of these two classes exhibits great potential for Xn detection. Molybdenum disulfide (MoS_2) and molybdenum trioxide (MoO_3) are very famous 2D materials with interesting electronic and optical properties.[38−40] MoS_2 exhibits semiconductor properties and possesses a direct band gap within the visible spectrum, rendering it highly promising for optoelectronic applications, whereas MoO_3 functions as a wide-band-gap semiconductor, which acts as an effective electron transfer mediator and exhibits excellent stability in aqueous environments, and has found utility in gas sensing and other domains. The amalgamation of MoS_2 and MoO_3 in a nanohybrid configuration yields stimulating properties and unveils possibilities for novel applications.[41] The MoS_2/MoO_3 nanohybrid has gained significant attention in diverse fields, including energy storage and electrocatalysis, among others. In energy storage applications, MoS_2/MoO_3 nanohybrids have been used as electrode materials for lithium-ion batteries and supercapacitors. The unique structure of the nanohybrid allows for efficient charge transfer and high capacitance, leading to improved performance compared to individual MoS_2 or MoO_3 materials.

Here, in this work, the heterostructure of a MoS_2/MoO_3 nanohybrid has been synthesized through a facile hydrothermal process, which enabled the large defect sites for Xn adsorption at the surface of the working electrode. No such work has been reported so far where the MoS_2/MoO_3 nanohybrid-based electrode has been used for Xn detection; therefore, the results hold great significance in this field. The high surface area of the nanohybrid allows for efficient enzyme adsorption, and the electronic properties of the MoS_2 component can be tuned to enhance the sensitivity and selectivity of the sensor to develop a proficient enzymatic biosensor.

EXPERIMENTAL METHOD

Synthesis of MoS_2 Nanosheets and the MoS_2/MoO_3 Nanocomposite. The synthesis of MoS_2 nanosheets was performed using the hydrothermal method. Initially, the precursor solution was prepared by dissolving ammonium molybdate and thioacetamide in deionized water in a 1:2 ratio to form the precursor solution. After stirring the precursor solution for 30 min, 0.05 g of cetyltrimethylammonium bromide (CTAB) was added as a surfactant to the solution, which helps to stabilize the nanosheets and prevent agglomeration. After the addition of the surfactant, the mixture was stirred for a certain amount of time to ensure that the surfactant was evenly distributed throughout the solution. Then, the solution was transferred to a Teflon-lined autoclave and heated at 200 °C for 36 h. After the hydrothermal process was complete, the resulting solution was cooled to room temperature and then centrifuged to separate the MoS_2 nanosheets from the solution. The nanosheets were then washed several times with deionized water and ethanol to remove any remaining impurities and surfactant. Finally, the MoS_2 nanosheets were dried under a vacuum to remove any remaining solvent. The MoS_2/MoO_3 nanocomposite was prepared by adding MoO_3 with ammonium molybdate in a certain ratio. The further process will be repeated as usual. The resulting powders were then characterized using various analytical techniques to determine their morphology and composition.

Electrophoretic Deposition (EPD) of MoS_2/MoO_3 Nanosheets. Before the electrodeposition of the MoS_2/MoO_3 nanocomposite over the indium tin oxide (ITO) surface, the electrode was cut into a 2 cm × 1 cm dimension and washed properly with an ethanol/deionized (DI) solution to remove the dirt and other impurities from its surface. The prepared nanosheets were deposited on a prehydrolyzed ITO electrode using the two-electrode EPD technique (Genetix, GX300C). Platinum was used as a counter electrode. In order to deposit films, the two electrodes were placed at a distance of 10 mm inside the sonicated solutions of the prepared samples in distilled water (DI) water. The deposition of films was achieved at 10 V for 25 s. Finally, the prepared films were kept in a vacuum oven for drying. Two deposited films have been retrieved, i.e., MoS_2/ITO and MoS_2/MoO_3/ITO, respectively.

Fabrication of the MoS_2/MoO_3/ITO Nanohybrid Biosensor. The immobilization was done by treating electrodes with 0.1% glutaraldehyde and allowed to rest for some time (1−2 h) at ambient temperature in order to activate the functional groups attached to the surface of the electrodes.

Afterward, phosphate-buffered saline (PBS; pH 7.4, 100 mM, 0.9% NaCl) was used to wash the electrodes, and then the covalent immobilization of XOD (20 μL) was done. The XOD/MoS_2/ITO and XOD/MoS_2/MoO_3/ITO electrodes were then kept in a moist condition at room temperature

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962−31971

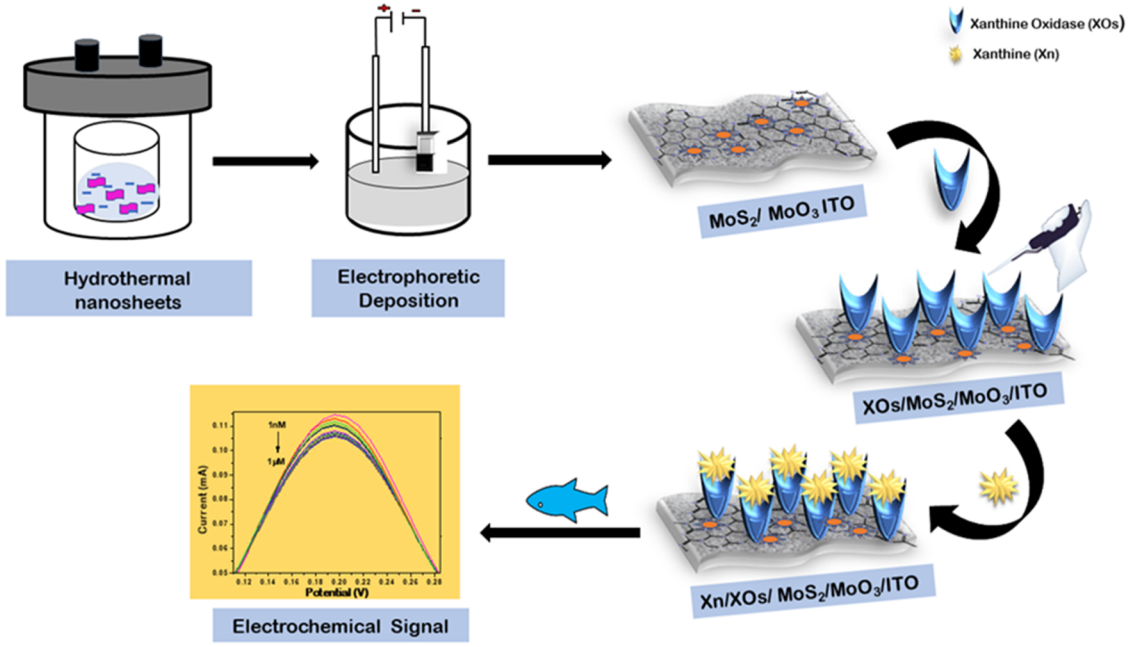


Figure 1. Scheme describing the fabrication of the MoS_2/MoO_3 nanosheet-based enzymatic biosensor.

overnight. The scheme for the electrode formation is shown in Figure 1.

RESULTS AND DISCUSSION

Structural and Morphological Studies. Raman spectroscopy serves as a potent technique for examining the structural and electronic characteristics of MoS_2/MoO_3, thereby offering valuable insights into the material's properties. Through the analysis of Raman spectra, crucial information regarding the nanocomposite's vibrational modes and crystalline structure was extracted, as illustrated in Figure 2.

The Raman spectrum of MoS_2/MoO_3 consists of several peaks corresponding to different vibrational modes. In MoS_2/MoO_3, the Raman spectrum of MoS_2 shows peaks related to the E_{1g} (396.28 cm^{-1}) and A_{1g} (424.42 cm^{-1}) modes of the crystal lattice, which are related to the in-plane and out-of-plane vibrations of the sulfur atoms, respectively (as shown in the inset). The Raman spectrum of MoO_3 shows peaks related to the symmetric and asymmetric stretching vibrations of the Mo−O bonds. The peaks in the spectrum include those at around 609 and 815 cm^{-1}, which are assigned to the B_{2g}, and E_g modes, respectively, depicting the occurrence of the $D2h$[16] space group. The occurrence of the MoS_2/MoO_3 nanohybrid was confirmed through the Raman analysis, and the difference between E_{1g} and A_{1g} (28.14 cm^{-1}) depicts the presence of a few layered composites.

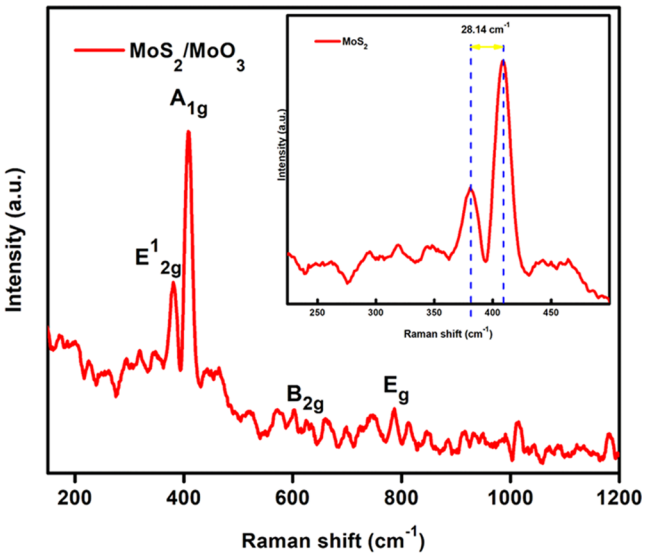


Figure 2. Raman spectrum of MoS_2/MoO_3 nanosheets; the inset shows the characteristic bands for MoS_2 nanosheets.

X-ray photoelectron spectroscopy (XPS) is a powerful surface analysis technique that provides information about the elemental and chemical composition, chemical bonding, and electronic structure of a material. The XPS has been recorded on ULVAC-PHI, PHI5000, which depicts the chemical states of the MoS_2/MoO_3 sample (shown in Figure 3a). Here, the binding energy of the core electrons of the constituent elements, including Mo, S, and O, would be observed. The binding energy of the Mo 3d, S 2p, and O 1s core levels has provided information about the oxidation state of the elements and the nature of the chemical bonding. The wide spectrum of MoS_2/MoO_3 nanosheets describes the presence of Mo, C, O, and S elements on the surface of the prepared nanoparticles. In Figure 3b, the two Mo bands observed at 228.55 and 231.76 eV are attributed to Mo $3d_{5/2}$ and Mo $d_{3/2}$, respectively, which signifies the occurrence of the Mo^{4+} state of molybdenum and their peak separation was measured to be 3.21 eV. The XPS spectra of the S 2p region, as shown in Figure 3c, represent two bands at 161.67 and 162.91 eV attributed to the occurrence of S $2p_{3/2}$ and S $2p_{1/2}$, respectively, which indicates the presence of the S^{2-} oxidation state of sulfur. In Figure 3d, the O 1s oxidation state at 531.65 eV is attributed to the occurrence of MoO_3. The XPS spectrum confirms the occurrence of the MoS_2/MoO_3 composites.

Figure 4 shows the field emission scanning electron microscopy (FESEM) images obtained using a JEOL JSM 6610LV instrument for MoS_2/MoO_3, which typically showed a

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962−31971

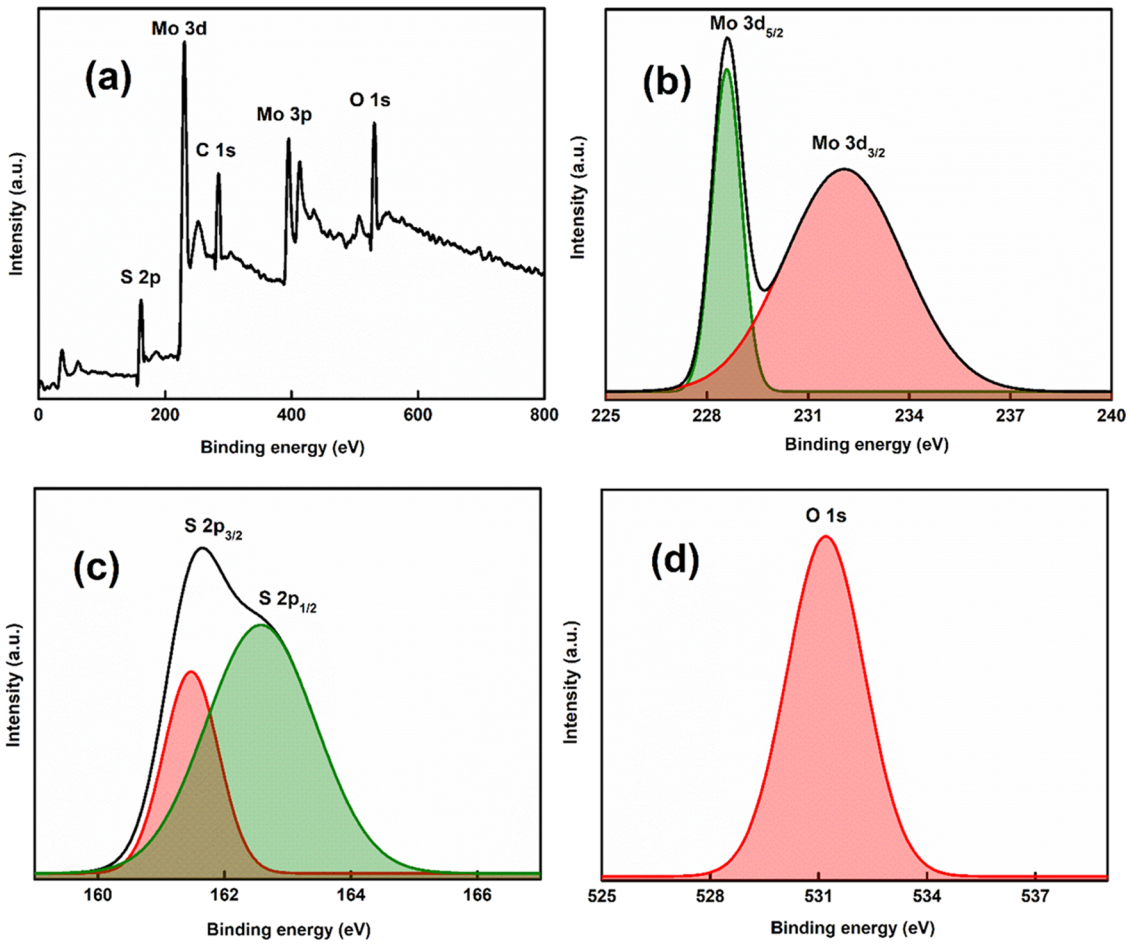


Figure 3. XPS spectra of (a) MoS_2/MoO_3 nanosheets, (b) XPS map of the Mo 3d element, (c) XPS map of the S 2s element, and (d) XPS map of O 1s.

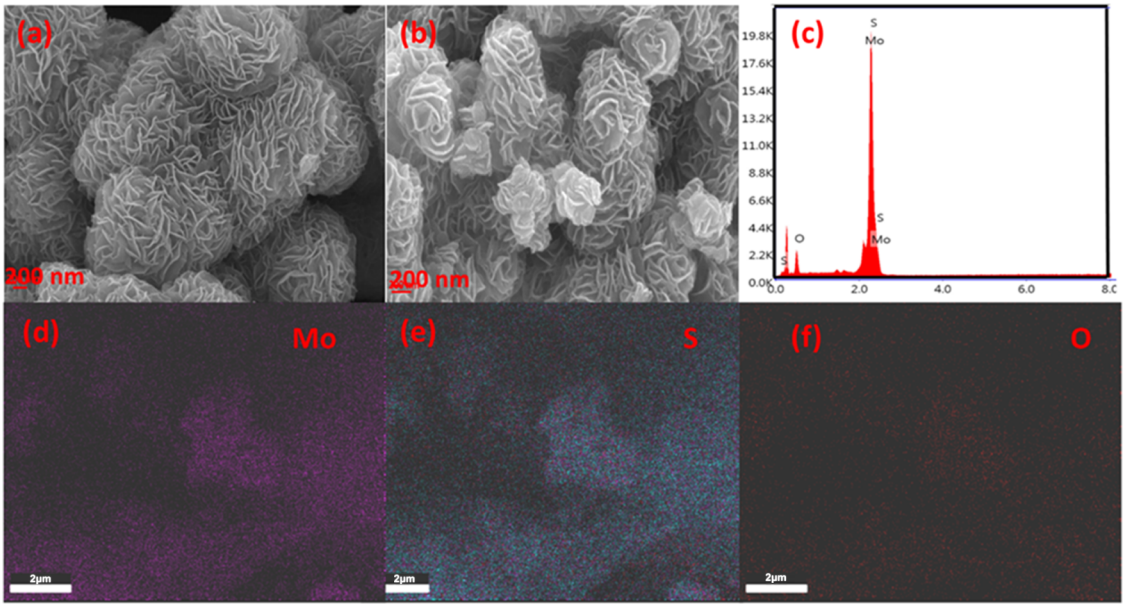


Figure 4. SEM images (a, b), EDX analysis (c), and elemental mapping of the components (d–f) for MoS_2/MoO_3 nanosheets.

layered structure consisting of MoS_2 and MoO_3 layers stacked on top of each other. The surface of the material appears rough and porous due to the presence of MoO_3 nanoparticles. The size and distribution of these nanoparticles were analyzed using FESEM images, which came out to be 400–600 nm (Figure 4a,b). Energy-dispersive X-ray spectroscopy (EDS) was used to analyze the elemental composition of the MoS_2/MoO_3 nanohybrid, which confirmed the formation of the composite

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962–31971

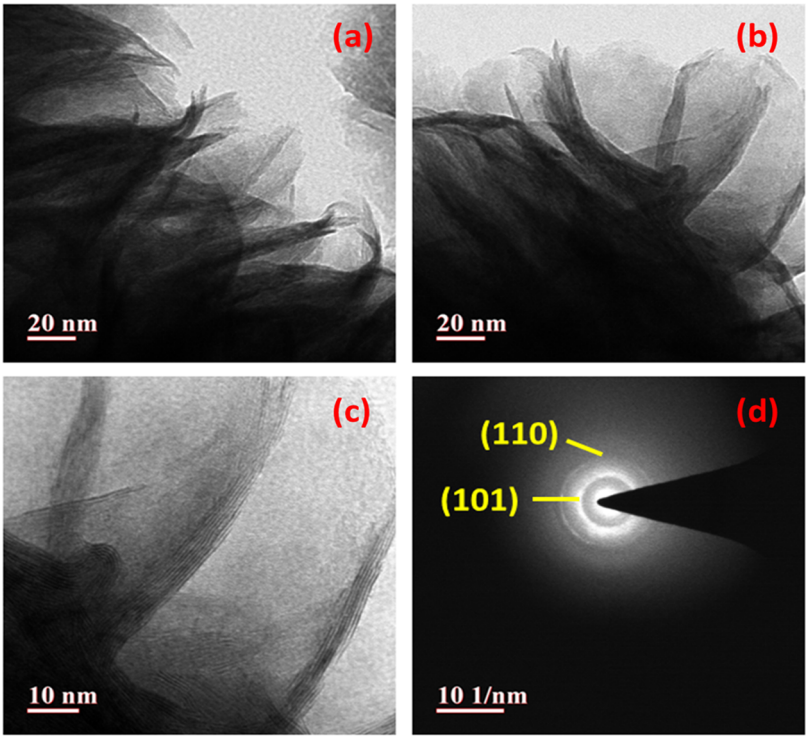


Figure 5. TEM images at different resolutions: (a–c) SAED patterns of the synthesized MoS_2/MoO_3 nanosheets.

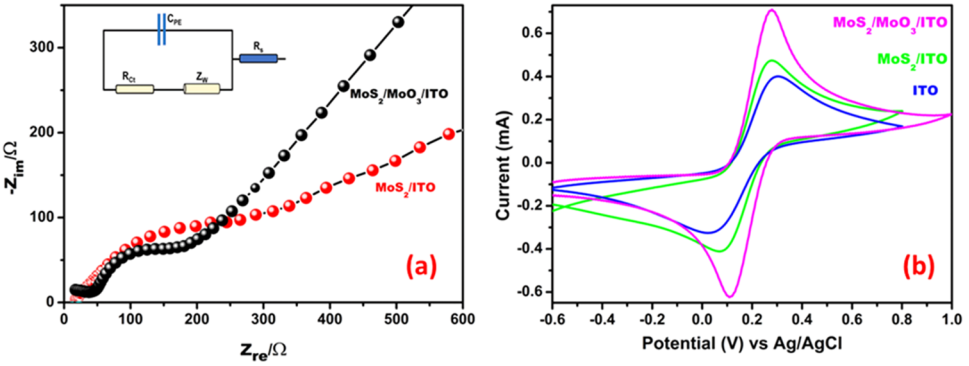


Figure 6. (a) EIS curves and (b) CV curves of prepared MoS_2/ITO and XOD/MoS_2/MoO_3/ITO electrodes.

(Figure 4c), and the elemental mapping was done to verify the homogeneity of constituent elements across the sample, in addition to XPS results (Figure 4d–f).

Transmission electron microscopy (TEM) analysis is done on a TECNAI 200 kV, which has provided valuable information about the morphology of the MoS_2/MoO_3 composite. The sample was imaged at various magnifications to observe the structure and morphology of the MoS_2/MoO_3 composite, as shown in Figure 5. The TEM images were recorded for the sample dispersed in DI water, loaded on a copper TEM grid with a carbon coating. The TEM images illustrate the layered structure of the composite stacked over one another. The figure illustrates the formation of a few layers counted from the edges of the sheets. It shows the presence of the 4–6 nanosheets in the prepared sample, which lead to high adsorption due to a greater surface-to-volume ratio. This confirms the results obtained from the Raman analysis of the composite material. The selected area electron diffraction (SAED) pattern shown in Figure 5d illustrates the high crystalline nature of the composite.

Electrochemical Studies of the Fabricated Electrodes. The electrochemical properties were observed on instrument Autolab potentiostat (Eco-Chemie, the Netherlands) for the prepared MoS_2/ITO and MoS_2/MoO_3/ITO heterostructure-based electrode. Electrochemical impedance spectroscopy (EIS; PBS, 5 mM $[Fe(CN)^6]^{3-/4-}$) within the frequency range of 0.01 to 10^5 Hz at a set potential of 0.01 V was carried out(Figure 6a).

EIS is particularly useful for investigating the interface between an electrode and an electrolyte solution, which is important for a wide range of applications. The data obtained from EIS can be analyzed through the Nyquist plot, which is a linear plot observed at lower frequencies, whereas at higher frequencies, a semicircular plot was recorded. This is attributed to the resistance because of the diffusion-controlled electron transfer and electron transfer limited process, respectively. The equivalent circuit has been displayed in the inset of Figure 6a. The charge transfer resistance (R_{ct}) of MoS_2/ITO and MoS_2/MoO_3/ITO was recorded to be 98.511 and 66.165 Ω, respectively.

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962–31971

From the obtained R_{ct} values, the exchange current per geometric unit area (i_o) and the apparent electron transfer rate constant (K_{app}) for MoS_2/ITO and MoS_2/MoO_3/ITO electrodes have been calculated using eqs 1 and 2

$$i_o = nRT/R_{ct}F \quad (1)$$

$$K_{app} = RT/n^2F^2AR_{ct}C \quad (2)$$

where R is the gas constant, F is the Faraday constant, T is the temperature in Kelvin, and A and C are the area of the electrode and the concentration of the redox probe, respectively. The i_o for the MoS_2/ITO electrode was found to be 26.06×10^{-5} A cm^{-2}, whereas, for the MoS_2/MoO_3/ITO electrode, the value of i_o was 38.80×10^{-5} A cm^{-2} respectively. Similarly, K_{app} for the MoS_2/ITO and MoS_2/MoO_3/ITO electrodes was calculated to be 2.14×10^{-6} and 3.19×10^{-6} cm s^{-1}, respectively. This increase in the values of i_o and K_{app} of the MoS_2/MoO_3/ITO electrode compared to the MoS_2/ITO electrode clearly describes the precise diffusion of the redox couple through the electrode interface.

The technique of cyclic voltammetry (CV) is widely utilized in electrochemistry to examine the redox characteristics of materials. In this study, the CV analysis of ITO (curve i), MoS_2/ITO (curve ii), and MoS_2/MoO_3/ITO (curve iii) electrodes was carried out by subjecting the working electrode to a voltage ramp ranging from −1 to 1 V while monitoring the resulting current (Figure 6b). There was a direct flow of electrons between the electrode and the ferro−ferri electrolyte, leading to an increase in the oxidation signal of MoS_2/MoO_3/ITO (curve iii). This finding suggests that the MoS_2/MoO_3/ITO electrode is a promising candidate for future biosensing applications. The charge transfer coefficient (α) and heterogeneous electron transfer coefficient (K_s) for MoS_2/MoO_3/ITO have been calculated to be 0.9073 and 0.0342, respectively, at a scan rate (v) of 50 mV/s (using Laviron's model eq 3).[47,48]

$$\ln(K_s) = \alpha \ln(1-\alpha) + (1-\alpha)\ln(\alpha) - \ln\left(\frac{RT}{nFv}\right) - \alpha(1-\alpha)\frac{nF\Delta E_p}{RT} \quad (3)$$

For the MoS_2/MoO_3/ITO (Figure 7a) electrode, the performance of the electrodes was investigated by varying the scan rate (10 to 150 mV/s). It has been observed that the oxidation peak current increases linearly with a positive shift in the peak potential. Both the anodic peak current (I_{pa}), as well as cathodic peak current (I_{pc}) increase linearly with increasing the scan rate; the following eqs 4 and 5 show the surface adsorption-controlled process kinetics for the MoS_2/MoO_3/ITO electrode (Figure 7b).

$$I_{pa}\ (\text{mA}) = 0.112 + 0.0538\sqrt{v}\ (\text{mV/s})$$
$$R = 0.996 \quad (4)$$

$$I_{pc}\ (\text{mA}) = -0.129 - 0.0384\sqrt{v}\ (\text{mV/s})$$
$$R = -0.993 \quad (5)$$

The anodic and cathodic peak potentials (E_{pa} and E_{pc}) also showed a linear relation with the natural logarithm of the scan rate (ln v). The corresponding linear regression equations eqs 6 and 7 have been given in Figure 7c.

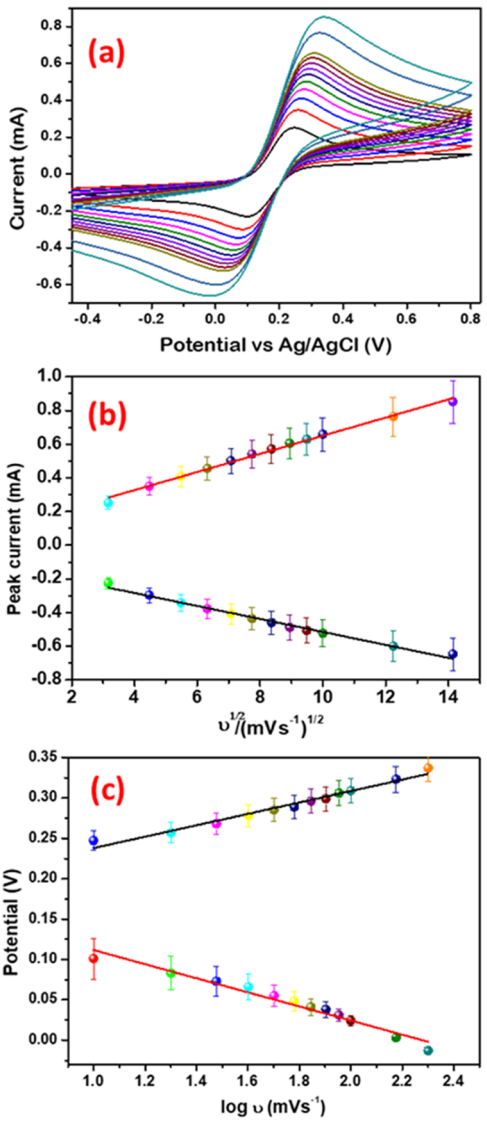


Figure 7. (a) CV at different scan rates, (b) anodic and cathodic peak current, and (c) anodic and cathodic peak voltage of the MoS_2/MoO_3/XOD/ITO electrode.

$$E_{pa}\ (\text{V}) = 0.0706 \ln\ln(v) + 0.1623;\ R = 0.9896 \quad (6)$$

$$E_{pc}\ (\text{V}) = -0.0872 \ln(v) + 0.1988$$
$$R = -0.9819 \quad (7)$$

Electrochemical Biosensing Response Studies of the XOD/MoS_2/MoO_3/ITO Electrode. Xanthine is typically detected through an enzymatic reaction catalyzed by xanthine oxidase (XOD) immobilized on the electrode surface. The MoS_2/MoO_3 nanocomposite can act as a mediator in the electron transfer process. The nanocomposite structure facilitates the efficient transfer of electrons between the immobilized enzyme (XOD) and xanthine, generating an electrochemical signal. To perform the biosensing studies, the XOD/MoS_2/MoO_3/ITO electrode was used to measure the electrochemical response under various concentrations of Xn (Figure 8a). The current response was recorded as the Xn concentration was increased from 1 nM to 1 μM. A linear decrease in current was observed with an increase in Xn concentration, indicating direct electron transfer between Xn

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962−31971

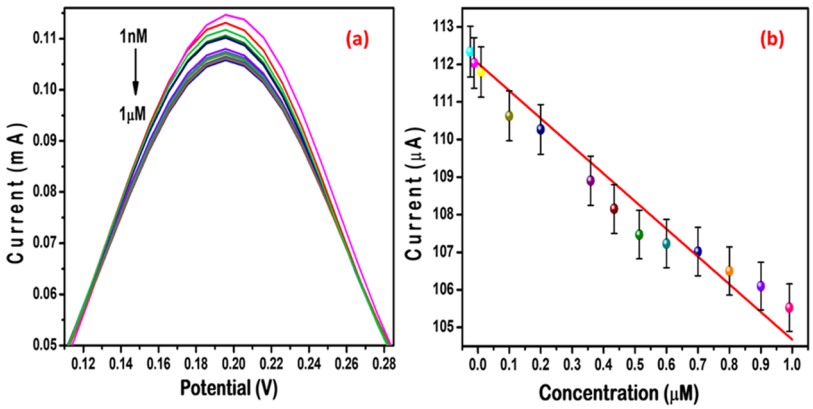


Figure 8. (a) DPV studies showing the current response of the MoS_2/MoO_3/XOD/ITO electrode and (b) current calibration plot with increasing concentration of Xn.

and the XOD/MoS_2/MoO_3/ITO electrode (Figure 8b). The biosensor's sensitivity toward Xn was measured using the regression equation $y = 112.034 - 7.357 \times 10^{-6}x$, where y represents the recorded peak current and x is the concentration of Xn, resulting in a sensitivity of 29.198 μA M^{-1} cm^{-2}. The limit of detection (LOD) and the limit of quantization (LOQ) were found using the equation 3.3 σ/S and (10 × LOD)/3.3, respectively, where σ is the standard deviation of the bioelectrode and S is sensitivity.[42,43] The calculated values for LOD and LOQ are 64 and 177 nM, respectively.

Real-Sample (Rohu Fish Extract) Validation of the Biosensor. Differential pulse voltammetry (DPV) was used to evaluate the practicality of the MoS_2/MoO_3/ITO biosensor for the direct detection of Xn in fish meat samples.

Xn was extracted from Rohu (*Labeo rohita*) fish meat using the method of Watanabe et al. The meat was collected from a nearby market, washed, and chopped finely, and 7 g of the meat was centrifuged in 15 mL of DI water at 4500 rpm. The surfactant was filtered, and the resulting solution was diluted 100 times to achieve the desired concentration. Electrochemical studies demonstrated a linear decrease in the peak current with an increase in Xn concentration from 0.01 to 0.5 μM (Figure 9), similar to synthetic Xn samples. The XOD/MoS_2/MoO_3/ITO biosensor was compared to other Xn-based sensors for various biosensing parameters (Table 1). The results showed that the biosensor exhibited a better response toward the primary detection of Xn in fish meat samples with a lower limit of detection (LOD).

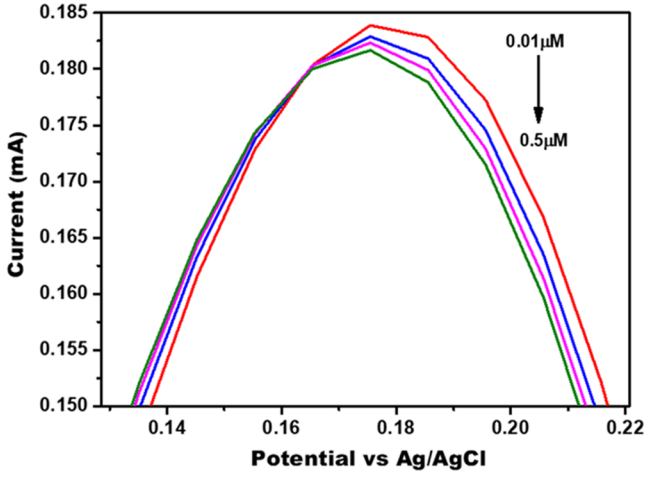


Figure 9. DPV studies of the real sample from fish showing the current response of the XOD/MoS_2/MoO_3/ITO electrode with increasing concentration of Xn.

Selectivity, Reproducibility, Reusability, and Shelf-Life Studies. The nanohybrid biosensor (XOD/MoS_2/MoO_3/ITO) was evaluated for cross-reactivity against interfering analytes (glucose, sodium benzoate, urea, uric acid, and ascorbic acid) having a 10-fold higher concentration than Xn.

No significant changes were observed (Figure 10a); the current remained stable for all other analytes represented as a mixture. The reproducibility of the XOD/MoS_2/MoO_3/ITO electrodes was confirmed by fabricating five similar electrodes under the same conditions (Figure 10b). The response of the fabricated biosensors to 1 nM of Xn ranged from 90 to 99%, with a relative standard deviation (RSD) value of 12% (n = 5). Stability tests were conducted to evaluate the sensor's reusability and shelf life, as shown in Figure 11. The single electrode was used 10 times after washing in saline, with no significant changes observed (Figure 11a).

The biosensor's shelf life was monitored by recording the DPV response toward Xn every week while being stored at 4 °C. After 5 weeks, the electrochemical response decreased from its original activity, indicating its stability over a long period (Figure 11b).

CONCLUSIONS

Detecting Xn is highly important in the food industry as it can signal the spoilage of fish meat in its early stages. Therefore, here, an electrochemical enzymatic biosensor has been fabricated for the primary detection of Xn in fish meat by employing the 2D MoS_2/MoO_3 nanosheets. The 2D arrangement of MoS_2 nanosheets provides a large active surface area, and the inclusion of MoO_3 acts as an electron mediator and enhances the electrochemical response of the electrode, as well as it is more cost-effective than previously used materials such as gold nanoparticles. The developed enzymatic sensor demonstrated a fast response time of 10 s and a detection limit of approx. 64 nanomoles, whereas the limit of quantization came out to be 177 nM. The accuracy of the biosensor was validated using real samples of xanthine extracted from rohu fish meat purchased from a local market, and similar trends were observed in the results for xanthine

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962−31971

Table 1. Comparison of the Performance of XOD/MoS_2/MoO_3/ITO with Other Biosensors for Xn Detection

s.no.	immobilization matrix	detection technique	linear range (μM)	LOD	refs
1.	Au-XOR/fMWCNT-PEDOT/GCE	DPV	0.1−10	5.45×10^{-2} μM	44
2.	XODNPs/Au	CV	0.01−1	0.01 μM	45
3.	XO/poly(L-asp)/MWCNT/GCE	DPV	0.001−0.004	3.5×10^{-4} μM	46
4.	XOD/ZnO/Ch/c-MWCNT/PANI/Pt	CV	0.1−100	0.1 μM	47
5.	Au-PPy	CV	0.4−100	0.4 μM	48
6.	XOD/c-MWCNT/PANI/Pt	CV	0.6−58	0.6 μM	49
7.	Ch/PPy/AuNPs/GCE	CV	1−200	0.25 μM	50
8.	PANI@TiO_2/ITO	DPV	1−100	0.1 μM	3
9.	poly-TTCA/Au	CV	5 − 0.001	1 μM	51
10.	XOD/MoS_2/MoO_3/ITO	DPV	0.001−1	64 nM	present work

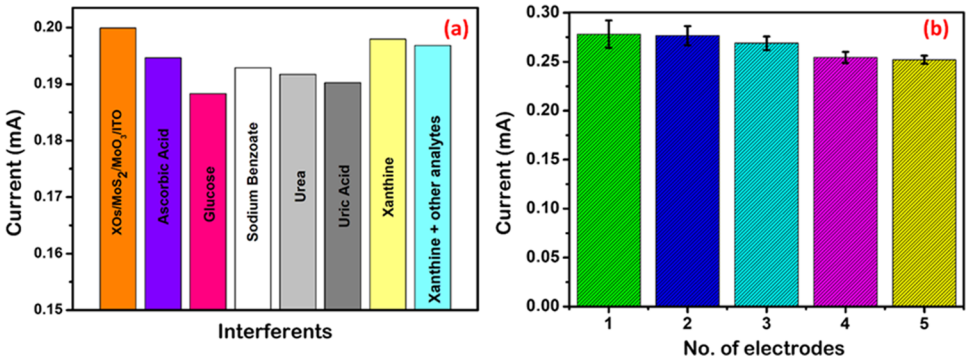


Figure 10. (a) Specificity and (b) reproducibility tests of the XOD/MoS_2/MoO_3/ITO electrode for Xn detection in the fish sample.

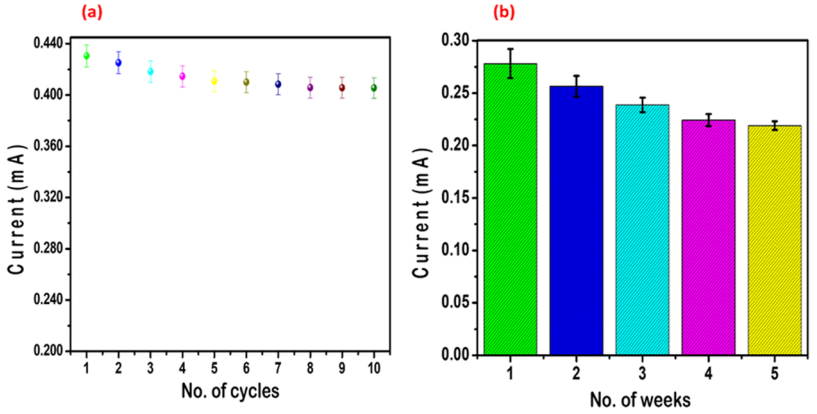


Figure 11. (a) Reusability tests conducted upto 10 cycles and (b) shelf-life study of upto 5 weeks of the fabricated enzymatic biosensor.

detection. The biosensor was found to be stable, specific, and reproducible. The proposed enzymatic biosensing platform based on the MoS_2/MoO_3 nanohybrid has been validated successfully for fish meat and exhibited its reliable applicability toward fish freshness parameters.

AUTHOR INFORMATION

Corresponding Author

Prateek Sharma − *GNIOT Institute of Professional Studies, Greater Noida Institute of Technology, Greater Noida, Uttar Pradesh 201310, India;* orcid.org/0000-0003-3366-8771; Email: prateeksharma1688@gmail.com, drprateek.bca@gniot.net.in

Authors

Deeksha Thakur − *Department of Applied Chemistry, Delhi Technological University, Delhi 110042, India;* orcid.org/0000-0003-4205-7139

Devendra Kumar − *Department of Applied Chemistry, Delhi Technological University, Delhi 110042, India;* orcid.org/0000-0001-9118-2070

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c03776

Author Contributions

P.S. contributed to conceptualization, data curation, formal analysis, investigation, methodology, and writing and editing—original draft. D.T. contributed to conceptualization, data curation, formal analysis, investigation, methodology, and draft editing. D.K. contributed to supervision, project administration, and draft editing.

Notes

The authors declare no competing financial interest.

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962−31971

■ ACKNOWLEDGMENTS

The authors would like to acknowledge the Greater Noida Institute of Technology, Delhi Technological University, and the University of Delhi for providing the appropriate infrastructure and facilities required for promoting this research work. D.T. acknowledges the DST-INSPIRE for the SRF Award.

■ REFERENCES

(1) Reza, K. K.; Singh, M. K.; Yadav, S. K.; Singh, J.; Agrawal, V. V.; Malhotra, B. D. Quantum Dots Based Platform for Application to Fish Freshness Biosensor. *Sens. Actuators, B* **2013**, *177*, 627−633.

(2) Sharma, N. K.; Monika; Kaushal, A.; Thakur, S.; Thakur, N.; Sheetal; Kumar, D.; Bhalla, T. C. Nanohybrid Electrochemical Enzyme Sensor for Xanthine Determination in Fish Samples. *3 Biotech* **2021**, *11*, No. 212.

(3) Thakur, D.; Pandey, C. M.; Kumar, D. Highly Sensitive Enzymatic Biosensor Based on Polyaniline-Wrapped Titanium Dioxide Nanohybrid for Fish Freshness Detection. *Appl. Biochem. Biotechnol.* **2022**, *194*, 3765−3778.

(4) Devi, R.; Yadav, S.; Nehra, R.; Yadav, S.; Pundir, C. S. Electrochemical Biosensor Based on Gold Coated Iron Nanoparticles/Chitosan Composite Bound Xanthine Oxidase for Detection of Xanthine in Fish Meat. *J. Food Eng.* **2013**, *115*, 207−214.

(5) Choi, H. K.; Liu, S.; Curhan, G. Intake of Purine-Rich Foods, Protein, and Dairy Products and Relationship to Serum Levels of Uric Acid: The Third National Health and Nutrition Examination Survey. *Arthritis Rheum.* **2005**, *52*, 283−289.

(6) Lawal, A. T.; Adeloju, S. B. Comparison of Enzyme Immobilisation Methods for Potentiometric Phosphate Biosensors. *Biosens. Bioelectron.* **2009**, *25*, 406−410.

(7) Ghanbari, K.; Nejabati, F. Ternary Nanocomposite-Based Reduced Graphene Oxide/Chitosan/Cr2O3 for the Simultaneous Determination of Dopamine, Uric Acid, Xanthine, and Hypoxanthine in Fish Meat. *Anal. Methods* **2020**, *12*, 1650−1661.

(8) Rezaeinasab, M.; Benvidi, A.; Gharaghani, S.; Abbasi, S.; Zare, H. R. Electrochemical Investigation of the Inhibition Effect of Carvacrol on Xanthine Oxidase Activity Merging with Theoretical Studies. *Process Biochem.* **2019**, *83*, 86−95.

(9) Torres, A. C.; Ghica, M. E.; Brett, C. M. A. Design of a New Hypoxanthine Biosensor: Xanthine Oxidase Modified Carbon Film and Multi-Walled Carbon Nanotube/Carbon Film Electrodes. *Anal. Bioanal. Chem.* **2013**, *405*, 3813−3822.

(10) Shih, M. H.; Lazo, M.; Liu, S. H.; Bonekamp, S.; Hernaez, R.; Clark, J. M. Association between Serum Uric Acid and Nonalcoholic Fatty Liver Disease in the US Population. *J. Formosan Med. Assoc.* **2015**, *114*, 314−320.

(11) Ono, T.; Tsuruta, R.; Fujita, M.; Aki, H. S.; Kutsuna, S.; Kawamura, Y.; Wakatsuki, J.; Aoki, T.; Kobayashi, C.; Kasaoka, S.; Maruyama, I.; Yuasa, M.; Maekawa, T. Xanthine Oxidase Is One of the Major Sources of Superoxide Anion Radicals in Blood after Reperfusion in Rats with Forebrain Ischemia/Reperfusion. *Brain Res.* **2009**, *1305*, 158−167.

(12) Davis, S. N.; Lastra-Gonzalez, G. Diabetes and Low Blood Sugar (Hypoglycemia). *J. Clin. Endocrinol. Metab.* **2008**, *93*, No. E2.

(13) Kim, K. Y.; Schumacher, H. R.; Hunsche, E.; Wertheimer, A. I.; Kong, S. X. A Literature Review of the Epidemiology and Treatment of Acute Gout. *Clin. Ther.* **2003**, *25*, 1593−1617.

(14) Kostić, D. A.; Dimitrijević, D. S.; Stojanović, G. S.; Palić, I. R.; Dordević, A. S.; Ickovski, J. D. Xanthine Oxidase: Isolation, Assays of Activity, and Inhibition. *J. Chem.* **2015**, *2015*, 1−8.

(15) Hlavay, J.; Haemmerli, S. D.; Guilbault, G. G. Fibre-Optic Biosensor for Hypoxanthine and Xanthine Based on a Chemiluminescence Reaction. *Biosens. Bioelectron.* **1994**, *9*, 189−195.

(16) Kito, M.; Tawa, R.; Takeshima, S.; Hirose, S. Fluorometric Determination of Hypoxanthine and Xanthine in Biological Fluids by High-Performance Liquid Chromatography Using Enzyme Reactors. *J. Chromatogr. B: Biomed. Sci. Appl.* **1983**, *278*, 35−42.

(17) Olojo, R. O.; Xia, R. H.; Abramson, J. J. Spectrophotometric and Fluorometric Assay of Superoxide Ion Using 4-Chloro-7-Nitrobenzo-2-Oxa-1,3-Diazole. *Anal. Biochem.* **2005**, *339*, 338−344.

(18) Berti, G.; Fossati, P.; Tarenghi, G.; Musitelli, C.; d'Eril, G. V. M. Enzymatic Colorimetric Method for the Determination of Inorganic Phosphorus in Serum and Urine. *Clin. Chem. Lab. Med.* **1988**, *26*, 399−404.

(19) A quantitative method for the analysis of xanthine alkaloids in Paullinia cupana (guarana) by capillary column gas chromatography | Semantic Scholar, 2022. https://www.semanticscholar.org/paper/A-quantitative-method-for-the-analysis-of-xanthine-Pagliarussi-Freitas/06172977e6d1a1eb6b0182532a9e71245a7f37da. (accessed November 03, 2022).

(20) Kock, R.; Delvoux, B.; Greiling, H. A High-Performance Liquid Chromatographie Method for the Determination of Hypoxanthine, Xanthine, Uric Acid and Allantoin in Serum. *Clin. Chem. Lab. Med.* **1993**, *31*, 303−310.

(21) Thakur, D.; Pandey, C. M.; Kumar, D. Graphitic Carbon Nitride-Wrapped Metal-Free PoPD-Based Biosensor for Xanthine Detection. *ACS Omega* **2022**, *8*, 2328−2336.

(22) Kalimuthu, P.; Leimkühler, S.; Bernhardt, P. V. Low-Potential Amperometric Enzyme Biosensor for Xanthine and Hypoxanthine. *Anal. Chem.* **2012**, *84*, 10359−10365.

(23) Wu, Y.; Deng, P.; Tian, Y.; Feng, J.; Xiao, J.; Li, J.; Liu, J.; Li, G.; He, Q. Simultaneous and Sensitive Determination of Ascorbic Acid, Dopamine and Uric Acid via an Electrochemical Sensor Based on PVP-Graphene Composite. *J. Nanobiotechnol.* **2020**, *18*, No. 112.

(24) He, Q.; Wu, Y.; Tian, Y.; Li, G.; Liu, J.; Deng, P.; Chen, D. Facile Electrochemical Sensor for Nanomolar Rutin Detection Based on Magnetite Nanoparticles and Reduced Graphene Oxide Decorated Electrode. *Nanomaterials* **2019**, *9*, No. 115.

(25) Wu, Y.; Deng, P.; Tian, Y.; Ding, Z.; Li, G.; Liu, J.; Zuberi, Z.; He, Q. Rapid Recognition and Determination of Tryptophan by Carbon Nanotubes and Molecularly Imprinted Polymer-Modified Glassy Carbon Electrode. *Bioelectrochemistry* **2020**, *131*, No. 107393.

(26) He, Q.; Tian, Y.; Wu, Y.; Liu, J.; Li, G.; Deng, P.; Chen, D. Facile and Ultrasensitive Determination of 4-Nitrophenol Based on Acetylene Black Paste and Graphene Hybrid Electrode. *Nanomaterials* **2019**, *9*, No. 429.

(27) He, Q.; Liu, J.; Liu, X.; Li, G.; Chen, D.; Deng, P.; Liang, J. Fabrication of Amine-Modified Magnetite-Electrochemically Reduced Graphene Oxide Nanocomposite Modified Glassy Carbon Electrode for Sensitive Dopamine Determination. *Nanomaterials* **2018**, *8*, No. 194.

(28) He, Q.; Liu, J.; Liu, X.; Li, G.; Chen, D.; Deng, P.; Liang, J. A Promising Sensing Platform toward Dopamine Using MnO2 Nanowires/Electro-Reduced Graphene Oxide Composites. *Electrochim. Acta* **2019**, *296*, 683−692.

(29) Zeng, R.; Wang, W.; Cai, G.; Huang, Z.; Tao, J.; Tang, D.; Zhu, C. Single-Atom Platinum Nanocatalyst-Improved Catalytic Efficiency with Enzyme-DNA Supermolecular Architectures. *Nano Energy* **2020**, *74*, No. 104931.

(30) Zeng, R.; Huang, Z.; Wang, Y.; Tang, D. Enzyme-Encapsulated DNA Hydrogel for Highly Efficient Electrochemical Sensing Glucose. *ChemElectroChem* **2020**, *7*, 1537−1541.

(31) Huang, L.; Cai, G.; Zeng, R.; Yu, Z.; Tang, D. Contactless Photoelectrochemical Biosensor Based on the Ultraviolet-Assisted Gas Sensing Interface of Three-Dimensional SnS2Nanosheets: From Mechanism Reveal to Practical Application. *Anal. Chem.* **2022**, *94*, 9487−9495.

(32) Lv, S.; Zhang, K.; Zhu, L.; Tang, D. Zif-8-Assisted Nayf4:Yb,Tm@zno Converter with Exonuclease Iii-Powered Dna Walker for near-Infrared Light Responsive Biosensor. *Anal. Chem.* **2020**, *92*, 1470−1476.

(33) Rahman, M. M.; Balkhoyor, H. B.; Asiri, A. M. Ultra-Sensitive Xanthine Sensor Development Based on Wet-Chemically Prepared Co/ZnO Nanoparticles. *Mater. Express* **2017**, *7*, 93−103.

(34) Rahman, M. M.; Marwani, H. M.; Algethami, F. K.; Asiri, A. M. Xanthine Sensor Development Based on ZnO−CNT, ZnO−CB,

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962−31971

ZnO−GO and ZnO Nanoparticles: An Electrochemical Approach. *New J. Chem.* **2017**, *41*, 6262−6271.
(35) Alam, M. M.; Asiri, A. M.; Uddin, M. T.; Islam, M. A.; Rahman, M. M. Wet-Chemically Prepared Low-Dimensional ZnO/Al 2 O 3 /Cr 2 O 3 Nanoparticles for Xanthine Sensor Development Using an Electrochemical Method. *RSC Adv.* **2018**, *8*, 12562−12572.
(36) Hussain, M. M.; Rahman, M. M.; Asiri, A. M.; Awual, M. R. Non-Enzymatic Simultaneous Detection of l -Glutamic Acid and Uric Acid Using Mesoporous Co 3 O 4 Nanosheets. *RSC Adv.* **2016**, *6*, 80511−80521.
(37) Alam, M. M.; Asiri, A. M.; Uddin, M. T.; Islam, M. A.; Awual, M. R.; Rahman, M. M. Detection of Uric Acid Based on Doped ZnO/Ag2O/Co3O4 Nanoparticle Loaded Glassy Carbon Electrode. *New J. Chem.* **2019**, *43*, 8651−8659.
(38) Sharma, P.; Singh, M. K.; Mehata, M. S. Sunlight-Driven MoS_2 Nanosheets Mediated Degradation of Dye (Crystal Violet) for Wastewater Treatment. *J. Mol. Struct.* **2022**, *1249*, No. 131651.
(39) Sharma, P.; Mehata, M. S. Colloidal MoS_2 Quantum Dots Based Optical Sensor for Detection of 2,4,6-TNP Explosive in an Aqueous Medium. *Opt. Mater.* **2020**, *100*, No. 109646.
(40) Sharma, P.; Mehata, M. S. Rapid Sensing of Lead Metal Ions in an Aqueous Medium by MoS_2 Quantum Dots Fluorescence Turn-Off. *Mater. Res. Bull.* **2020**, *131*, No. 110978.
(41) Singh, S.; Deb, J.; Sarkar, U.; Sharma, S. MoS_2/MoO_3 Nanocomposite for Selective NH3Detection in a Humid Environment. *ACS Sustainable Chem. Eng.* **2021**, *9*, 7328−7340.
(42) Yu, Z.; Qiu, C.; Huang, L.; Gao, Y.; Tang, D. Microelectromechanical Microsystems-Supported Photothermal Immunoassay for Point-of-Care Testing of Aflatoxin B1 in Foodstuff. *Anal. Chem.* **2023**, *95*, 4212−4219.
(43) Zeng, R.; Qiu, M.; Wan, Q.; Huang, Z.; Liu, X.; Tang, D.; Knopp, D. Smartphone-Based Electrochemical Immunoassay for Point-of-Care Detection of SARS-CoV-2 Nucleocapsid Protein. *Anal. Chem.* **2022**, *94*, 15155−15161.
(44) Sen, S.; Sarkar, P. An Interference-Free New Xanthine Biosensor Based on Immobilized Enzyme-Nanogold Conjugate on Carbon Nanotube Doped Poly(3,4-Ethylenedioxythiophene) Composite Film. *Int. J. Biol. Macromol.* **2022**, *199*, 275−286.
(45) Joon, A.; Ahlawat, J.; Aggarwal, V.; Jaiwal, R.; Pundir, C. S. An Improved Amperometric Determination of Xanthine with Xanthine Oxidase Nanoparticles for Testing of Fish Meat Freshness. *Sens. Bio-Sens. Res.* **2021**, *33*, No. 100437.
(46) Yazdanparast, S.; Benvidi, A.; Abbasi, S.; Rezaeinasab, M. Enzyme-Based Ultrasensitive Electrochemical Biosensor Using Poly(l-Aspartic Acid)/MWCNT Bio-Nanocomposite for Xanthine Detection: A Meat Freshness Marker. *Microchem. J.* **2019**, *149*, No. 104000.
(47) Devi, R.; Yadav, S.; Pundir, C. S. Amperometric Determination of Xanthine in Fish Meat by Zinc Oxide Nanoparticle/Chitosan/Multiwalled Carbon Nanotube/Polyaniline Composite Film Bound Xanthine Oxidase. *Analyst* **2012**, *137*, 754−759.
(48) Devi, R.; Yadav, S.; Pundir, C. S. Au-Colloids-Polypyrrole Nanocomposite Film Based Xanthine Biosensor. *Colloids Surf., A* **2012**, *394*, 38−45.
(49) Devi, R.; Yadav, S.; Pundir, C. S. Electrochemical Detection of Xanthine in Fish Meat by Xanthine Oxidase Immobilized on Carboxylated Multiwalled Carbon Nanotubes/Polyaniline Composite Film. *Biochem. Eng. J.* **2011**, *58*−*59*, 148−153.
(50) Dervisevic, M.; Dervisevic, E.; Çevik, E.; Şenel, M. Novel Electrochemical Xanthine Biosensor Based on Chitosan−Polypyrrole−Gold Nanoparticles Hybrid Bio-Nanocomposite Platform. *J. Food Drug Anal.* **2017**, *25*, 510−519.
(51) Rahman, M. A.; Won, M. S.; Shim, Y. B. Xanthine Sensors Based on Anodic and Cathodic Detection of Enzymatically Generated Hydrogen Peroxide. *Electroanalysis* **2007**, *19*, 631−637.

https://doi.org/10.1021/acsomega.3c03776
ACS Omega 2023, 8, 31962−31971

I.J. Image, Graphics and Signal Processing, **2023, 4, 1-18**
Published Online on August 8, 2023 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2023.04.01

Object tracking via a Novel Parametric Decisions based RGB-Thermal Fusion

Satbir Singh*
Dr B R Ambedkar National Institute of Technology/ Centre for Artificial Intelligence, Jalandhar, 144088, India
E-mail: satbirs.ai@nitj.ac.in
ORCID iD: https://orcid.org/0000-0002-6263-2092
*Corresponding Author

Arun Khosla
Dr B R Ambedkar National Institute of Technology/ Centre for Artificial Intelligence, Jalandhar, 144088, India
ORCID iD: https://orcid.org/0000-0001-8571-7614

Rajiv Kapoor
Delhi Technological University, Delhi, 110042, India
ORCID iD: https://orcid.org/0000-0003-3020-1455

Received: 08 September, 2022; Revised: 02 November, 2022; Accepted: 25 January, 2023; Published: 08 August, 2023

Abstract: The thermo- visual fusion based tracking has been deployed for overcoming the shortcomings of alone vision-based object tracking. The assistance from both domains should be wisely merged so that it should result in a useful practice for object tracking. Several techniques had been developed recently to implement a brilliant fusion, but this undeveloped field still inhibits many unsolved challenges. The proposed method aims at increasing the effectiveness of tracking by bi-modal fusion with the introduction of a new set of rules based upon the parameters generated from the decision of individual modality trackers. This practice helps to achieve output by only a single run of the fusion process in every frame. The method also proposes to use minimal information from individual trackers in normal conditions and incorporates the use of supplementary information from imageries merely in case of diverse working conditions. This procedure, in turn, lessens the computations and hence reduces time to process. The experiments performed on well-known publically available datasets show the advantages of the proposed method over the individual visual domain tracking and other existing states of the art fusion techniques.

Index Terms: Particle filter, Object tracking, Decision level fusion, visible-thermal amalgamation.

1. Introduction

Object tracking using sensor fusion of thermal data and the coloured video is a more efficient approach as compared to tracking in isolated individual domains. Visible camera mainly depends on the colour content of the object and surroundings. Thermal / IR camera only captures the thermal image of the object. The thermal image is exclusively related to the temperature value of the object. High-temperature objects possess precious intensity values, whereas low-temperature objects do not possess significant intensity in a thermal image. This information is beneficial for surveillance applications. Moreover, in the case of human target tracking, background illumination variations and inconsistent cloth texture can act as difficulties in vision-based monitoring, but the thermal domain imagery is unaffected from it.

On the other hand, the visible camera has its advantage by having more variations in the data acquired through the colour content information. It is also helpful when a similar temperature object comes in contact with one another, and it is difficult to differentiate their thermal distributions. Hence, during disguised conditions and under the presence of challenging complications, both can complement each other effectively. Since both the domains lie in the image plane, it is easiest to calibrate the information obtained from these cross domains. The following subsection introduces the readers with the previous work done in this motivational direction.

1.1. Related Work

Various research studies have been exercised in this sensor fusion technology. Face recognition inhibiting thermal

This work is open access and licensed under the Creative Commons CC BY License.

and visible sensor fusion using background subtraction had been proposed earlier in literature [1-4]. It has also been actively utilized in robotics based vision applications [5-7]. Besides this, the use of visible and thermal imagery in fusion has also been hosted in spacecraft proximity operations [8] to overcome the effect of different lightening conditions in the orbital environment. [9,10] have advanced the field of image fusion of multisensory data for the improved computer vision applications. More can be explored in the study of [11,12].

Notably, the use of said bi-modal information has been increased for object tracking under challenging conditions of vision in recent times. Nowadays researchers have been trying to advance this overwhelming field by the application of intense methods such as convolutional neural networks [13] that can automatically extract the deep features. A two-stream convolution net was formed independently for each domain and the fusion was achieved by developing a fusion net. Despite an exciting approach, it is to be noted that it is dependent on a rigorous training process through the use of image net database. For obtaining the instantaneous tracks, the process cannot track unless it has been prepared for feature extraction through a complicated process.

Extending a primary particle filter [14] method for adaptive fusion at particle level [15] was a good idea for tracking under similar background conditions. It introduced an additional parameter in fusion to avoid the camouflaging effect. However, the method lacked in several key aspects such as having a long processing time and repetition of fusion for every particle per each frame. Also, if the total number of particles is increased initially, then computational efficiency is affected and time to process rises severely.

A basic approach to tracking by fusion [16] involved multiplication of the individual specialists that used cues from thermal as well as visible imagery to form their opinions. However, simple multiplication only serves the purpose of video sequences with no veiling conditions. In case of false tracking in one mode of imagery, it may lead to overall false result due to the absence of procedure to verify this.

[17] presented a method in which the fusion was performed single time per frame due to the execution of some weighted fusion rules. In this method, individual tracking was performed in two different algorithms. Particle filter was incorporated for vision/colour and correlation based template matching was applied for thermal counterpart. However, this method also lacked in two significant aspects. First, any one domain is over weighted (>1) and other is under weighted (<1) in case of fusion rule 2 and fusion rule 3 in the paper, which are called for most of the time. The over dependency on one domain is based upon the only difference in maximum likelihood obtained from individual trackers dealing in foreground tracking only, which may be either false or true under challenging conditions. Second, this decision based strategy does not provide any correction in output due to the position of the least weighted imagery domain output, which at long distance may deviate itself and thus affects the location of final estimate through the presented fusion.

For the knowledge of readers, different methods that incorporated thermo-visible fusion based tracking are listed under Table1. The proposed work inspires to resolve the above-said issues of fusion, and the detailed contributions and motivation for the research are mentioned in the following subsections.

1.2. Motivation for the Work

A significant motivation for the research work lays in working out a technique that may be able to make a robust track outcome estimate by fusion of the decision made by individual trackers. So, the repetition of the same fusion rule per single step of tracking procedure should be avoided. The fact is that various existing approaches in literature employ the fusion either at the pixel level of the image or at the feature level of individual particles present in the particle filter. Therefore, both approaches in the categories mentioned above seem to be inefficient at sinking the computing time because the same fusion strategy is executed several times for finding out a single frame track outcome. If it is a pixel level fusion, then the fusion process repetition rate becomes equal to the size of the image, i.e., no. of pixels in the image.

Moreover, in the case of the most commonly used particle filter based tracking, fusion is applied for P turns per single tracking outcome. Here P denotes the total no. of particles used for tracking. So, computation cost related to the fusion portion of the tracking algorithm increases with an increase in the number of particles used. Moreover, some methods in literature [14,18] involve the role of background both in the thermal and visible imagery at almost every frame. This calculation of background region, its feature extraction and then the estimation of fusion parameters are repeated for each particle per single frame analysis, which makes it an ample time consuming task. So, the purpose of this research is to formulate a method that can amalgamate the role of both modalities with temporal efficiency and still provide an accurate track estimate. Another inspiration comes from the fact that significant work is done given likelihood adaptation of individual imagery, but work needs to be done in an adaptation of the role of the single state position estimate in a decision-based fusion framework.

1.3. Contribution of work

The proposed method stays with the use of single feature based particle filter tracking in individual domains but introduces an intelligent fusion framework that can overcome the limitations of either of the individual imagery trackers and results into a credible track estimate. The critical contributions of the proposed research work are mentioned in the following points:

i. The proposed work introduces an efficient design framework to obtain track estimates by using a fusion of

thermal and visible imageries. This presented fusion strategy makes it possible that the execution of the fusion step for imageries is not done more than once per single frame as compared to many times per frame execution in previous algorithms. This type of execution, in turn, saves much computational processing.

ii. Second, it presents a selective dealing perspective framework for the tracker. The tracker can either chose to perform simple fusion procedure under normal conditions or can decide to take the supplement help from imageries for making fusion decision only in the presence of diverse challenging circumstances.

iii. A formula is worked out for adjustment of positional coordinates of fused track output after decisive fusion process depending on the outcome provided by the output parameters of individual trackers. This helps in remedying the side effect of hazards in state estimate due to less voted tracker.

iv. A new occlusion treatment is also proposed that is based upon the introduced fusion process. Here, the occlusion detection is the result of the merged opinion from both the imageries.

The organization of the paper is as: Section 2 provides a description of work methodology adopted containing the details of the individual domain particle filter tracking followed by the detailed explanation of the designed fusion strategy along with all rules and parts. The latter part of this section mentions the procedure used while handling occlusions. The following section provides the experimental part along with the discussion on the results acquired covering the qualitative and quantitative aspects. Section 4 contains the concluding remarks and the future possibilities of the performed research.

2. Work Methodology Description

The following section describes the complete details of the work procedure followed for this technique of tracking. The whole process is divided into two main steps. The first subsection provides the insight of the particle filter algorithm incorporated for individual domain tracking. The following subsection provides details of the implementation of the proposed fusion algorithm to obtain the final track estimate.

2.1. Particle Filter Tracking

The implemented particle filter algorithm can be understood appropriately by breaking down the procedure into five stages. Each stage (except initialization of particle filter) was repeated for every frame of video. The explanations of these are as following:

2.1.1. Initializing Particle filter

Initially, on the availability of the target template, its features for reference were generated and fed into the system to act as a reference during the tracking process. To begin with the working of the particle filter, particles of the particle filter were distributed randomly across the scene frame around the target. Each particle was described using a state vector corresponding to $[X, Y, A, B, V_x, V_y]$. In this, X, Y correspond to starting coordinates of the rectangle shaped particle. A, B denote the width and height of the rectangle respectively. V_x and V_y are the velocities, with which, the particles move through the sequence in X and Y direction respectively.

2.1.2. State updating using motion model

After the start, circulated particles states were modified in every frame using a state updating procedure. A linear update matrix, U was used to make the increment in each proceeding frame. Equation (1) describes the state update process for the particles:

$$s_t^p = U \times s_{t-1}^p + G_t \tag{1}$$

Here, s_t^p stands for the p^{th} particle state at time instant $t = t$ and U is taken as (6X6) Identity Matrix, and $G_t \overset{\text{def}}{=} N(0, Q)$. Q is a zero mean Gaussian process with varianceσ_g, whose value is chosen as 0.25 during the experiments.

2.1.3. Observational features from particles

After creation and then updating particles, the next step was to extract information from the state of particles in every frame. Since the fusion algorithm uses two significant aspects: one is weight value and another the positional coordinates of the tracking outcome; therefore, two main features were extracted from every particle. The weight of the particle was awarded based upon a similarity measure calculation.

Table 1. Previous work related to combined Thermo-Visual object tracking

Reference	Algorithms used for tracking	Fusion Strategy	Application
C.Li [13]	Deep Convolutional Networks	Fusion net	RGB-Thermal Tracking
M. Talha et al. [14]	Particle Filter	Adaptive Foreground-Background relation based fusion	Tracking in camouflaging conditions
O'Conaire et al. [15]	Mean Shift tracker	Product of Expert Beliefs	Object tracking
G. Xiao et al. [16]	Particle filter, template matching	Weighted Fusion Rules	Object tracking
Stolkin et al. [19]	Particle Filter	Bayesian Fusion	Human Detection and Tracking
E. Fendri [20]	Background modelling	Logical AND between the binary mask regions in separate imageries	Moving object detection
S. Singh [21]	Particle Filter	Granular computing based fusion	Object tracking in datasets with difficult vision conditions
Y. Niu [22]	Frame difference based background subtraction	Fusion based on target region segmentation	Unmanned Aerial Vehicles (UAV)
S. Singh [23]	Template matching	Correlation based weighted fusion	Object tracking
C. Li [24]	Bayesian Filtering	Adaptive fusion using Collaborative Sparse representation	Tracking in Grayscale-Thermal framework
C. Li [25]	Bayesian Filtering	Laplacian sparse representation	Tracking in Grayscale-Thermal framework
H. P. Liu [26]	Particle Filter Tracking	Using Min operation on Sparse representation	OCTVBS dataset
Palmerini et al. [8]	Kalman Filter	Extended Kalman Filter	Spacecraft Proximity Tracking under light varying conditions

The use of Bhattacharya similarity measure was incorporated to associate the colour distributions of the template of the target to be tracked and the regions estimated from the state of different particles. First, colour histogram with eight bins for each R, G, and B components was found for both the template chosen as well as the region of the image associated with the individual particle in the visible domain. However, the thermal domain made use of only 1D grey intensity histogram. If b denotes the number of bins in the histogram, then similarity measure BS^p for particle p was given by (2) as:

$$BS^p = \sum_{k=1}^{b} \sqrt{T(k) \times R(k)} \tag{2}$$

Here T denotes the template RGB histogram and R stands for the particle region histogram. However, the distances were transformed to fit in the non-linear probabilistic robustness using (3) to weigh the individual particles.

$$w^p = \frac{e^{\{(BS^p)^2/2\sigma^2\}}}{\sqrt{2\pi\sigma}} \tag{3}$$

In the above equation, w^p stands for weight of p^{th} particle and σ is the variance parameter, whose value is taken as 0.01 throughout the experimentation. And the current state of the particle was marked as s_t^p, which contained the value of instantaneous location, dimensions and velocities of the individual particle.

2.1.4. State Estimation

The final state estimate was obtained by the average of the states obtained from each particle as shown in (4), i.e.

$$S_t = \frac{1}{N} \sum_{p=1}^{N} s_t^p \tag{4}$$

Also, the final combined likelihood parameter for the particle filter track at time t was found in similar manner and given by (5).

$$W_t = \frac{1}{N} \sum_{p=1}^{N} w_t^p \tag{5}$$

In the above equation, N denotes the total number of particles taken for tracking that were 50 for all the experimentation done in this research work. Whereas S_t, W_t stand for the combined weight and state estimate for the particle filter output for a single frame track procedure.

2.1.5. Resampling

Resampling process has been carried out to replace the particles that obtained lower likelihood with other particles in the particle sample for the next iteration. This process has been repeated in every frame to optimize the particle filter efficiency by filtering out the least required samples. The proposed algorithm exchanged the unworkable particles with S_t, which is the obtained mean state of the particles.

The individual track features were generated for the visible and thermal imagery counterpart separately and are denoted as W_t^V, S_t^V and W_t^T, S_t^T respectively. With the completion of the features extraction from the individual tracker, the final outcome was achieved using a new fusion technique, which has been presented in the following section of the paper.

2.2. Fusion Strategy

After obtaining the bimodal track information, selection of a suitable fusion rule was made to obtain track estimate. The fusion process mainly uses two critical statistics from the individual domain tracks to formulate its decision on the final track output estimate. The scheme, in general, can be illustrated using the following diagram in Fig. 1.

The fusion algorithm selects different fusion rules based upon the value of some observed parameters. First, the parameter 'Difference in Weights DIW_t has been calculated (given by (6)). Based upon DIW value, it selected one of two cases.

$$DIW_t = |W_t^V - W_t^T| \tag{6}$$

If DIW_t value was found to be higher than a regularly updated threshold value wd_t, then the sensor mode having the more significant weight was chosen as the correct tracking source after validating it through a background-foreground similarity check. For every frame, the value of wd_t was automatically calculated using (7).

$$wd_t = \frac{1}{2}\left(\frac{W_t^V + W_t^T}{2}\right) \tag{7}$$

The above said validation process in turn, helps to rectify the decision in case of presence of camouflaged backgrounds. Had the higher weighing sensor been tracking deceptively due to disguising conditions, it would have shown high background similarity value and the result would have been discarded. The details of the implementation have been provided in section 2.2.1 of the paper.

In next case, if DIW_t was found to be less than wd_t, that suggests both domains have comparable confidence level in their output track estimate. Hence, an added criteria was used to obtain an error free decision. For overcoming this situation, the section procedure now looked upon for the spatial information, which was established based on founding a new parameter 'Distance in States', DIS_t. If (X_t^V, Y_t^V) and (X_t^T, Y_t^T) were the initial coordinates in 2D space of region of track outcomes in a single sensory domain, then DIS_t was found as the Euclidean distance between these points in image space and is defined in (8) as:

$$DIS_t = \sqrt{(X_t^V - X_t^T)^2 + (Y_t^V - Y_t^T)^2} \tag{8}$$

Further, the state difference threshold value was modelled to make spatial information based decisions. Value of space difference threshold was chosen in reference to dimensions of the object being tracked. Value of sd was found using (9) which helped in revealing whether the difference was significant or not.

$$sd = \frac{1}{2}^2\sqrt{\left(\frac{A}{2}\right)^2 + \left(\frac{B}{2}\right)^2} \tag{9}$$

As mentioned earlier, A, B are the initial width and height of the object to be tracked.

Now, if value of DIS_t calculated was also minor, then it could be decided that estimated states were pointing towards a similar or nearby region in image space with alike confidence. Hence, both trackers resembled to an identical track estimate. So, the required output was found ready without using any Background validations that in turn saved lot of computational cost through a simple Foreground fusion framework detailed in section 2.2.2.

However, if DIS_t was significant (higher than value of), then it would have signified that although the weights obtained from individual trackers were similar, but these were indicating to two different locations. So, again background information was incorporated to aid the handicapped lone foreground detail. Section 2.2.3 describes the fusion for this case. The complete procedure of fusion strategy is as follows:

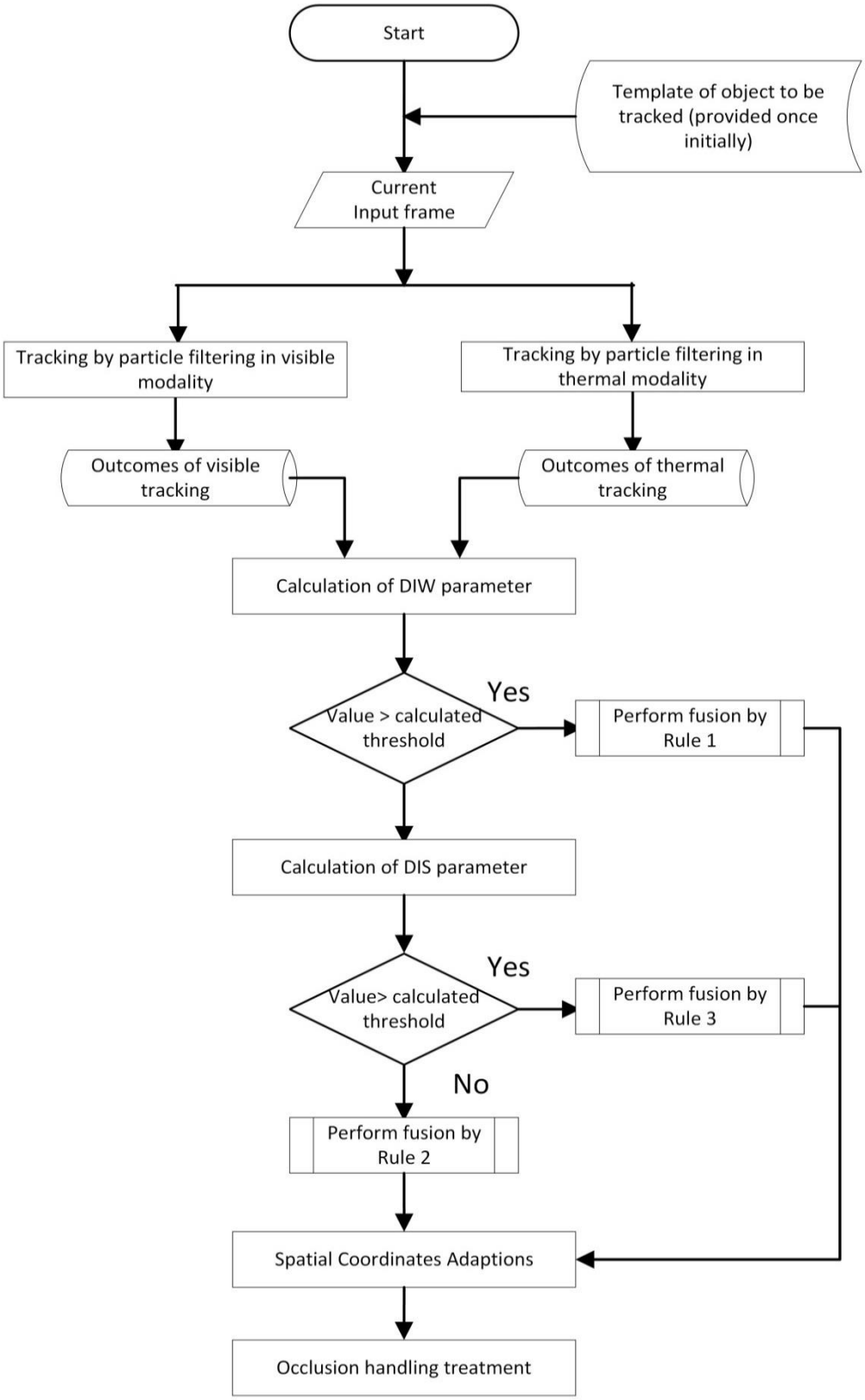


Fig. 1. Block representation of the work-flow procedure

2.2.1. Fusion Rule 1: $DIW_t > wd_t$ (Average weight obtained from one imaging modality was very high and obtained from other was too low).

In this case, due to very low likelihood indication given by one modality in self's decision, its opinion was not taken into consideration and the modality showing a superior likelihood of track was chosen as a sole leader for providing the track estimate. Further the sole dependency was authenticated only after passing the foreground–background validation process. Different techniques are available in the literature for finding the background region and foreground region relation. Our method adopted the local background model implementation of [17] to determine the background histogram of the state found.

If, the Bhattacharya similarity between reference target histogram and background region histogram found using (2) came out to be less than 0.5 (scale of 0-1), the following procedure for fusion was adopted:

i. Append that imagery's weight equal to unity, whose likelihood average weight was found to be far more than the other confidence diminished imagery.
ii. Make the other imagery's weight equal to zero.
iii. After changing the weights of individual imagery to either 0 or 1, the fusion output was calculated as:

$$S_t^F = \delta_t^V \times (S_t^V) + \delta_t^T \times (S_t^T) \tag{10}$$

Here δ_t^V and δ_t^T stand for modified weight values under the above said situation. The variable S_t^F denotes the output of fusion at time t after application of suitable fusion rule. The notation used for the fused output remains same throughout this paper.

Else, the situation was treated as no track in the current frame because, one tracker was having very low confidence in tracking and the other with a superior confidence was tracking wrongly either due to a deceiving condition or presence of similar background in its image signatures.

2.2.2. Fusion Rule 2: $DIW_t < wd_t$ and $DIS_t < sd$ (Average likelihood obtained from both imaging modalities and the location of track estimates from these were in close approximation).

Since the average confidence obtained from both imageries was nearby and also the output track locations pointed by each imagery resided closely in this case; therefore, a conjunctive consensus could be easily reached by use of a simple calculative procedure detailed in following steps:

i. Append Value of Foreground Adaptive Measure, γ, was found using (11) as:

$$\gamma = \frac{w_t^V}{w_t^V + w_t^T} \tag{11}$$

ii. The fused state estimate was updated according to (12).

$$S_t^F = \gamma \times (S_t^V) + (1 - \gamma) \times (S_t^T) \tag{12}$$

Hence, background calculations could be avoided for these situations that lessen the time consumption of the overall track process. The following subsection provides the implementation details of the last case of fusion.

2.2.3. Fusion Rule 3: $DIW_t < wd_t$ and $DIS_t > sd$ (Likelihood of both imaging modalities were found to be nearby but individual state estimate differed significantly).

For this different situation, the extra information that helped dissipate the ambiguity for the fusion was once again the local background conditions of the estimated track state outcome. The background details were found with the help of [17] as mentioned in subsection 2.2.1. Now, a reliable way to weigh the particles can be formulated in the following manner:

i. Value of adaptive background measure, β_t, was calculated based upon the similarities of backgrounds with the object reference template and was found using (13) as:

$$\beta_t = \frac{B_t^T}{B_t^T + B_t^V} \tag{13}$$

Where, B_t^T and B_t^V were the calculated Bhattacharyya similarity measures between the histograms of target object template and the background Regions of the states estimated in thermal and visible modality respectively. Both of these were found using (2).

ii. After obtaining a background measure, an effective way to provide confidence measures to the individual modes was adopted. This modelling is shown by (14) and (15):

$$\alpha_t = \frac{\beta_t . W_t^V}{\beta_t . W_t^V + (1-\beta_t) . W_t^T} \tag{14}$$

$$S_t^F = \alpha_t \times S_t^V + (1 - \alpha_t) \times S_t^T \tag{15}$$

Here, α_t is the parameter representing the weight adaptation parameter resulted from the introduction of the local background details of individual domain imagery. Rest of the parameters is the same as defined earlier in the paper.

The above rule indicates that the imaging modality that exhibited less similarity in its background region was multiplied with a larger weight factor as compared to the other one.

2.3. Improvement in Location of Track Estimate After Fusion

After updating the weights, it was necessary to cross-examine the relative spatial differences between the individual state estimates, as this could also degrade the performance of tracker. If one tracker is tracking correctly and has got a better confidence measure after above fusion process, but still the other tracker having even a small confidence measure could bring a significant shift in the correct track position. If it was pointing to a tracking estimate very far from the current position of the original target. So, despite adaptation in weights, the final track estimate might diverge without making an adaption in locations provided by individual trackers. Changes in states were made using the help of parameter relative to the weighting of individual trackers. First, W_{min} and W_{max} were found using (16) and (17) respectively.

$$W_{min} = \min(W_t^V, W_t^T) \tag{16}$$

$$W_{max} = \max(W_t^V, W_t^T) \tag{17}$$

After this, a factor was found for relative distance adjustments using the following relation presented in (18).

$$D = sd * \frac{W_{min}}{W_{max}} \tag{18}$$

After modification in weights, a check was made that the X and Y coordinates of the Track estimate region of least weighted domain differed (either less or more) from the X and Y coordinates of the other domain (having higher weight) by more than D pixels. If this condition existed, then a corresponding reduction or increment was made in the value of small weighted imagery's state coordinates so that their changed value had at most D pixels difference to remedy the unwanted shift. Otherwise, the values were kept unchanged.

2.4. Occlusion Rule

The method took the help of both the imaging modalities to detect the situation of occlusion in tracking. In the case of occlusion, the object or its parts hide behind another object present in the scene. So, the likelihood value suddenly gets decreased on the insertion of occluding. The proposed algorithm checked the similarity of the final track estimate in every frame with original track template available. This check was performed in both imagery domains concerned.

Moreover, the similarity measures obtained from both the domains were weighted with respect to their weights obtained after fusion. Now, the combined likeness degree was calculated using (19) as:

$$OS_t^F = W_t^V \times (OS_t^V) + W_t^T \times (OS_t^T) \tag{19}$$

Here, OS_t^V represents the Bhattacharyya measure between histogram densities for region of fused track estimate and the template histogram in the visible colour domain and OS_t^T is the corresponding counterpart in the thermal image. The values of these have been found using (19). If the value OS_t^F appeared to be higher than a threshold value for similarity, routine tracking process was continued. But if occlusion was detected due to a degraded similarity index, then algorithm moved towards occlusion handling procedure.

Once the occlusion was detected, then the algorithm searched for the robust presence of the object with the help of background subtraction using dynamic matrix factorization [27]. Once it was detected that the object comes out of occlusion, then the template model was updated by using the new conditions of the object, and the particles were redistributed around its location. The further routine tracking process continued till there is no occurrence of next occlude.

3. Experimental Results and Discussion

To verify the performance and scrutinize it with other states of the art trackers, we used MATLAB software on an I3 processor. The video sequences used along with the challenging conditions contained in these are written in the form of the following points:

i. OCTBVS Dataset Sequence 2 (Similarity in Backgrounds and visible camouflaging).
ii. OCTBVS Dataset Sequence 4 (Occlusion, Clutter)
iii. OCTBVS Dataset Sequence 6 (Presence of similar appearance objects, thermal camouflaging)
iv. INO Snow Parking Dataset (multiple alike objects, same appearance backgrounds, object rotation and most importantly effect of clutters).

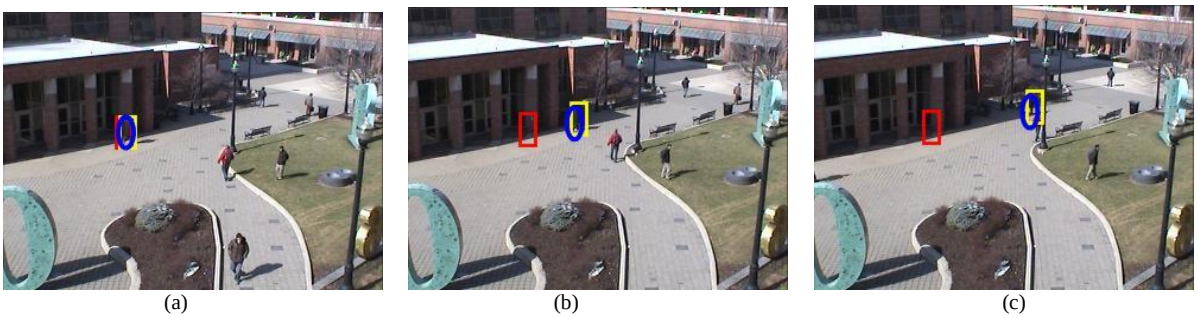
Experiments were performed and the results obtained were compared with 4 other algorithms viz. a) Colour Particle Filter, CPF [14], (b) Continuously adaptive data fusion of colour and gradients. CADF [18], (c) Adaptive fusion of thermal and Visible, AFTV [15], and (d) Tracking before fusion of infrared and visible, TBF [17]. The complete discussion on the results obtained has been narrated by observing the qualitative traits as well as the comparative statistical calculations from the performed experimentation.

3.1. Qualitative Demonstration of the utility of algorithm

The quality of a fusion algorithm can be judged from the fact that it should overcome the hurdles due to several conditions and unaffectedly chose the best outcome possible. Its performance should remain robust. This fact has been illustrated with the help of two video sequences:

3.1.1. OCTBVS Data Sequence 2

The advantage of fusion of the thermal modality in aid to the visible tracking can be demonstrated by Fig. 2 and Fig. 3. Here, the object moves across a colour camouflaged background conditions. It can be seen that there are dark spaces between the pillars of the building and also a tree present on the right side of the view that is making difficult for target judgment in the visible domain.



(a) (b) (c)

Fig. 2. Track outcomes shown on visible image view of seq. 2. Individual colour based tracking (red rectangle), Individual thermal imagery tracking (yellow rectangle) and fusion based tracking using the proposed algorithm (blue ellipse). (a) Frame no. 16, (b) Frame no. 89, (c) Frame 193

(a) (b) (c)

Fig. 3. Thermal Image counterpart of Fig. 2 in the sequence 2 of OSU dataset. (a) Frame no. 16, (b) Frame no. 89, (c) Frame 193

Whereas the thermal signatures robustly mark the presence of the object as compared to the visible captures due to the temperature difference between the human object and the background. It is interesting to note that when both the trackers are tracking accurately, the algorithm does not indulge in any complex computations by avoiding the application of any testing criteria for decisions. However, as soon as the visible particle filter starts estimating an erroneous trajectory, then the final fused track estimate remains with the accurately tracking domain due to the re-arrangements of the weighing criteria and proposes fusion state adjustments.

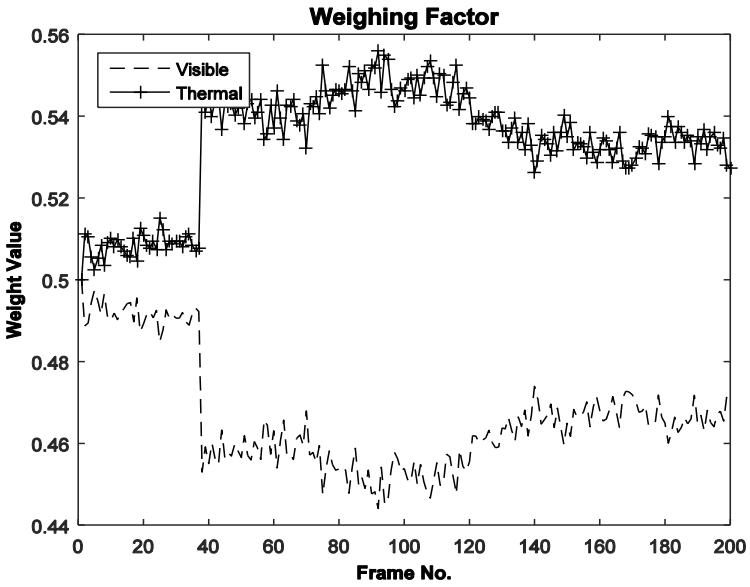


Fig. 4. Wight variation during sequence 2 videos

The fact can also be supported by Fig. 4, which shows that the weight of wrongly tracking visible domain starts decreasing and the fusion algorithm starts favouring the right tracking thermal domain after about 25 frames.

3.1.2. OCTBVS Data Sequence 4

The illustration shown in Fig. 5 and Fig. 6 suggests that the performance quality of the algorithm is not dependent solely on single imagery. The output during different frames of sequence 4 does not get affected when the thermal tracker tracks to incorrect output. Here rather than following the thermal imagery (which was earlier providing better estimates in the previous dataset), now the algorithm tracks by weighing in favour of the accurate tracking visible imagery. So, the fusion approach depreciates the role of thermal imagery in case of it providing false estimates.

(a) (b) (c)

Fig. 5. Track outcomes shown on visible image view of seq. 4. Individual colour based tracking (red rectangle), Individual thermal imagery tracking (yellow rectangle) and fusion based tracking using the proposed algorithm (blue ellipse). (a) Frame no. 45, (b) Frame no. 81, (c) Frame 153

(a) (b) (c)

Fig. 6. Thermal Image counterpart of Fig. 5 in the sequence 2 of OSU dataset. (a) Frame no. 45, (b) Frame no. 81, (c) Frame 153

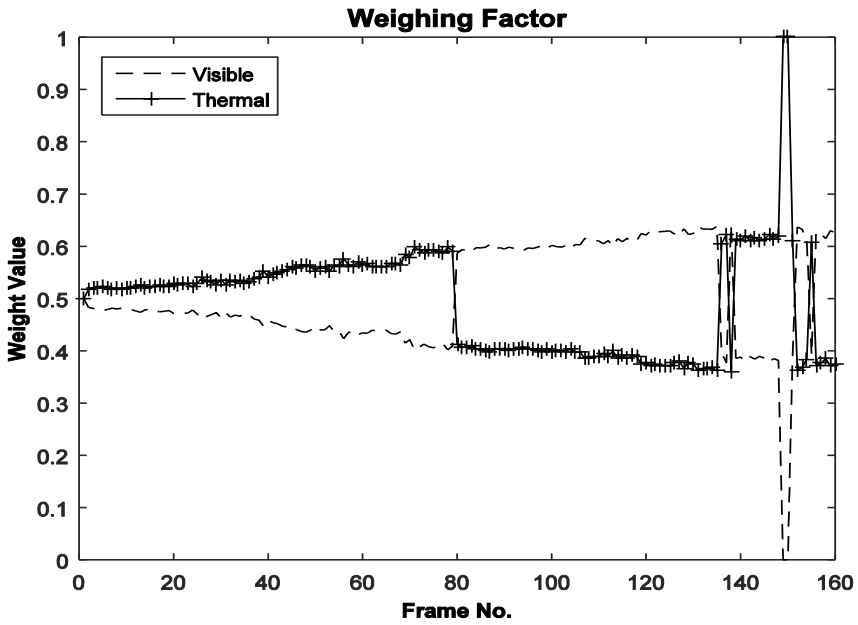


Fig. 7. Weight variation during sequence 4 videos

The same can also be illustrated through Fig. 7, which shows weight values provided to individual imageries after the fusion process. It is interesting to mention that under normal conditions till frame no. 78, the proposed technique finds the track estimate using simple weighted fusion as mentioned by fusion rule number 2 in the paper. This estimation used simple calculations. However, when a new person appeared near to the tracked target, there appeared some difference in opinions of individual trackers. Consequently, the algorithm shifted to fusion rule 3 and started to adjust weights more in favour of accurately tracking visible imagery.

3.2. Quantitative Comparisons with other methods

The accuracy was judged while comparing the experimental results with the ground truth positions of the object throughout the video sequence. This comparison was quantified in terms of two parameters Root Square Tracking Error and the statistical F-measure calculated with the help of precision and recall values. The significance of these parameters is that the value of one resembles the closeness in the position of the target tracked with the actual position of the object and the other represents the bounding box overlap efficiency of the results. The combined diversion from the ground truth positional values in both axes was represented in the form of Root Square Tracking Error, $RSTE_t$, whose value can be found using (20).

$$RSTE_t = \sqrt{\left(G_t(x) - S_t^F(x)\right)^2 + \left(G_t(y) - S_t^F(y)\right)^2} \tag{20}$$

In above equation, $G_t(x)$ and $G_t(y)$ are the X and Y coordinate of starting point of the actual target state respectively, whereas, $S_t^F(x)$ and $S_t^F(y)$ represent the x and y coordinate of starting point of the final track estimate found through applied fusion algorithm.

The next parameter F-measure is found by first finding the precision and recall values and then substituting these in (23). The precision (found using (22)) is defined as the ratio of an area covering intersection the ground truth bounding box with the estimated track bounding box to the area covered by the bounding box of the estimated track state. Whereas, recall can be found using (21), i.e., by the ratio of an area covering intersection the ground truth bounding box with the estimated track bounding box to the area covered by the original bounding box enclosing the object.

$$R_t = G_t(area) \cap S_t^F(area)/G_t(area)\ ) \tag{21}$$

$$P_t = G_t(area) \cap S_t^F(area)/S_t^F(area) \tag{22}$$

$$F_t = 2 \times P_t \times R_t/(P_t + R_t) \tag{23}$$

Here, P_t and G_t mark the current frame precision and recall values respectively. Whereas F_t is the statistical F-measure of the obtained track estimate.

3.2.1. OCTBVS Data Sequence 2

This sequence of OSU dataset contains a human pedestrian moving past a colour similarity background. It has been observed that as soon as the object passes the building walls after frame number 60 and is left only with black background (similar to the colour of dress the pedestrian is wearing), the ordinary colour-based particle filter turns out to be misguided and remains attached to the black background instead of tracking the real target. This effect is illustrated by the high deviation in the error curve shown in Fig. 8

Also, the continuously adaptive data fusion model of gradient and colour fusion remains in the proper track search till mid-way but loses the actual trajectory as it starts weighing more in favour of the colour domain which is already providing false track. Now, coming to the approach of thermal and visible fusion in [17], the track estimates are at some distance apart from the original position of the object consistently. This situation arises because though this fusion method improves the weights for the tracker by weighing more in favour of the thermal imagery, it does not take care of the positional coordinate arrangement. The larger the offshoot distance of false sensing domain, more deviation is present as observed in Fig. 8 and Fig. 9.

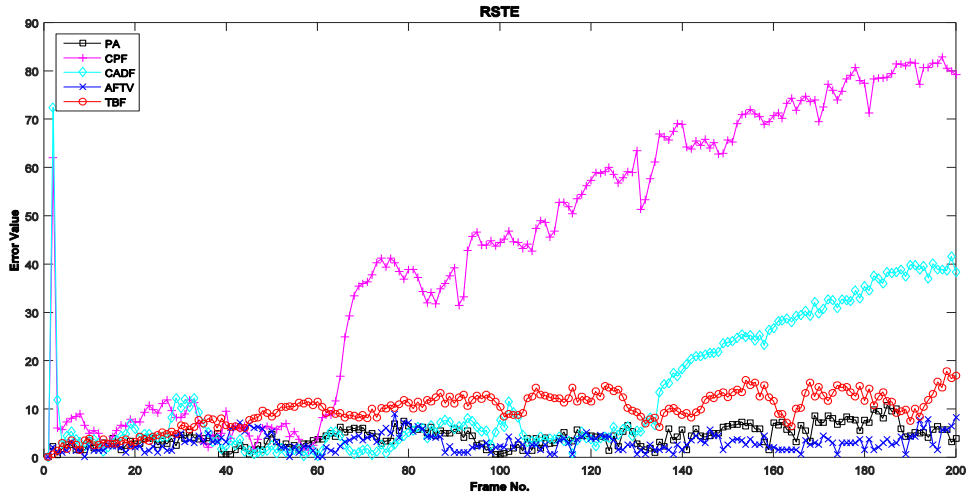


Fig. 8. RSTE plot for Sequence2

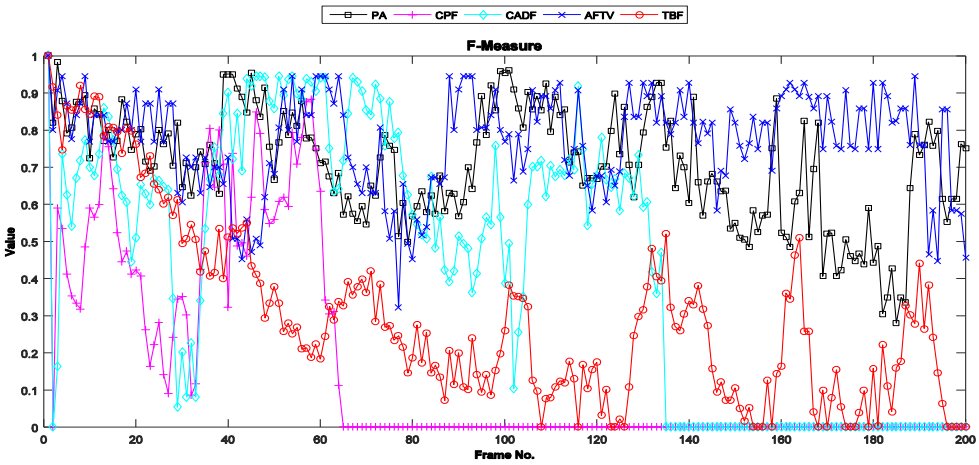


Fig. 9. F- Measure plot for Sequence2

The adaptive fusion of thermal and visible domain by [15] is providing correct weighting by adaptive background fusion model to weigh in favour of the thermal domain signatures of the view. However, this method is not computationally effective as it calculates a background model for every particle in every frame and then performs a similarity comparison for each of them. On the other side, the proposed method takes care of the track estimate by providing parameters both for weights updation as well as track state position updation. As indicated by Fig. 8 and Fig. 9, although the particle filter in colour domain is tracking falsely after frame 60, the thermal domain tracker keeps providing accurate track outcome. Accordingly, the proposed algorithm starts weighing more in favour of thermal domain tracker. Besides, a more substantial drift in location by the result of a false tracking domain has been encountered by the algorithm.

3.2.2. INO Dataset Snow Parking Sequence

The tracking in the snow parking dataset contained many challenging situations such as the presence of multiple alike objects, same colour backgrounds, object rotation and severe effect of clutters in the scene inside both imagery domains. The illustration is provided in Fig. 10 and Fig. 11.

It is notable to see that few other algorithms tend to lose their way as the moving target stops on the path for few frames and it interacts with a similar appearance pedestrian and rotates his motion direction several times during his communication. After frame 770, the object comes across a car that acts as a similar background object. Here rest of all trackers except the individual colour-based particle filter track correctly. Near to frame 780, the target comes in contact with a similar pedestrian that acts as clutter in both imageries, and hence many algorithms tend to hinder away from the original object.

This hazard has been indicated by an increase in the RSTE value (Fig. 12) and a decrease in the F-Measure value (Fig. 13). The AFTV algorithm also suffers ambiguities in the track estimate during frame interval of 790-860, under which the object rotates several times in various directions and comes back to original motion direction only after frame 870.

The TBF algorithm suffers during the presence of clutter and starts tracking correctly afterwards. The proposed algorithm performs a reasonable job throughout the whole frame interval due to amendments made both in weighing factors and the spatial positions.

(a) (b) (c) (d)

Fig. 10. Track outcomes of proposed algorithm shown on visible image view of SPDS video sequence (a) Frame no. 782, (b) Frame no. 796, (c) Frame 811, (d) Frame no.880

(a) (b) (c) (d)

Fig. 11. Thermal Image counterpart of Fig.10 in the sequence 2 of OSU dataset. (a) Frame no. 28, (b) Frame no. 60, (c) Frame 88, (d) Frame no.145

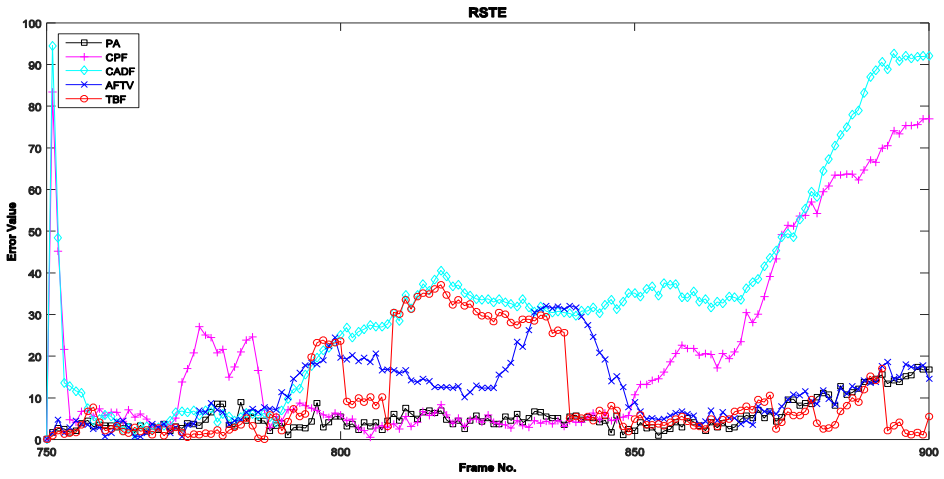


Fig. 12. RSTE plot for SPDS

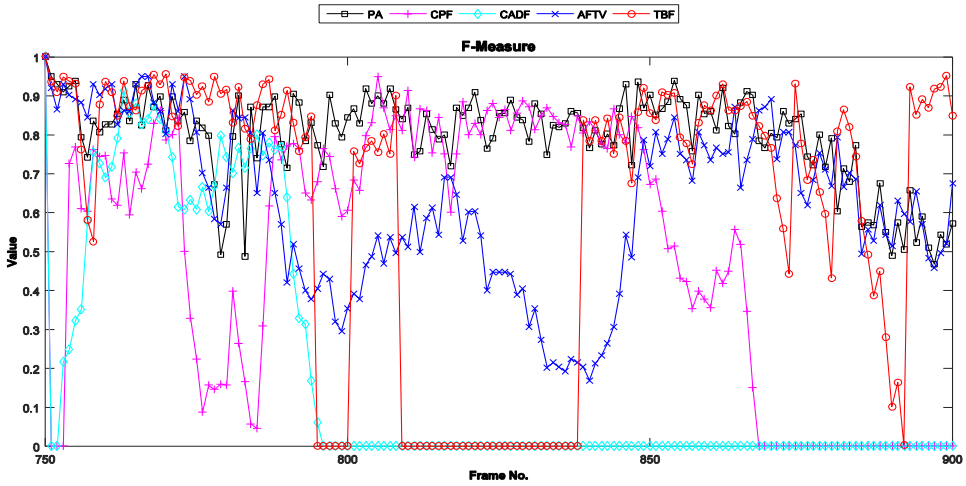


Fig. 13. F-Measure plot for SPDS

3.2.3. Sequence 6 Dataset

The scenes of this dataset contain challenging problems for a visual tracker like changing background conditions, the presence of similar feature objects, and also a similarity in the thermal domain. The object is initially in front of a white coloured car with black bottom front, and a black appearing mirror, but then its motion is traversed against empty space followed by passing a similar coloured pillar that also has it's lower half-size comparable to the dimensions of the object to be tracked. This situation engraves a challenging task, which is shown in Fig. 14 and Fig. 15.

(a) (b) (c) (d)

Fig. 14. Track outcomes of proposed algorithm shown on visible image of seq. 6. (a) Frame no. 28, (b) Frame no. 60, (c) Frame 88, (d) Frame no.145

(a) (b) (c) (d)

Fig. 15. Thermal Image counterpart of Fig.14 in the sequence 6 of OSU dataset. (a) Frame no. 28, (b) Frame no. 60, (c) Frame 88, (d) Frame no.145

Further, the presence of various coloured clutter objects such as black dustbin, windows of the building, etc. makes it tough for tracking using a lonely colour based tracker. The thermal image of the scene also faces a background camouflage in the initial part of the sequence that can be seen through Fig. 14(a).

No trackers except for two algorithms produced an accurate track estimate for this dataset containing various difficulties. The correct tracking was only achieved till nearby frame no. 40. The CADF was working based on the fusion of HOG and colour feature. However, it favoured the colour cue (not tracking accurately) as illustrated by the CPF curve in Fig. 15 due to more significant deviations in the colour likelihood of particle filter. The resultant is the start of ramp increase in the RTSE and decrease in F-Measure value (Fig. 16, Fig. 17). Though TBF favours mainly in the thermal domain from the start, it tracks below par due to thermal camouflaging present the individual thermal tracker. Moreover, with no adjustment in the final state positions, the fusion deviates a lot from the actual position of the target.

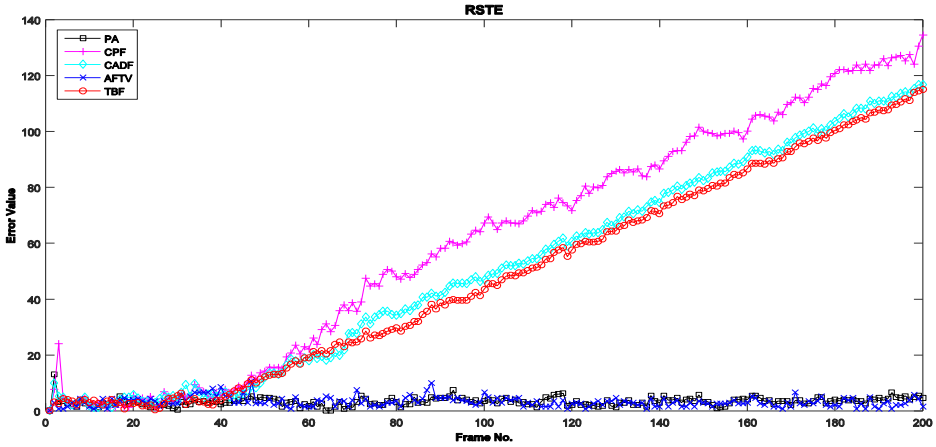


Fig. 16. RSTE Plot for Sequence 6 video

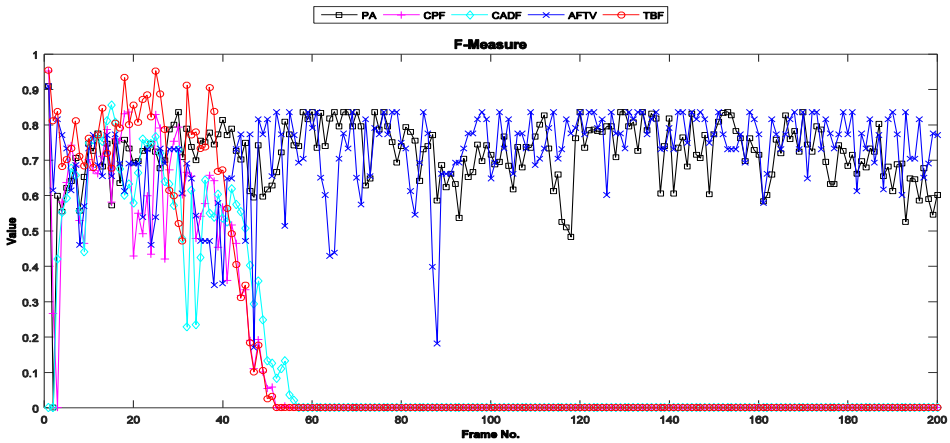


Fig. 17. F- Measure Plot for Sequence 6 video

The results have also been compared via a tabular representation providing a clear view of the performance of the different algorithms. Table 2 provides the time to process analysis, whereas Table 3 presents the RSTE comparison and F-measure statistics have been illustrated using Table 4. It is clear from Table 2 that the proposed algorithm only lacks behind the basic particle filter in timing analysis, but takes less time in comparison to other fusion schemes.

RSTE for various algorithms value was found using (23) for each frame. The average error has been tabulated in Table 3. All the results are averaged over five runs of experiments due to the random nature of the particle filter incorporated in each method. The last column of the table shows the overall average performance.

The best has been written in bold text and italics has been used for the second position. It is interesting to note that though the proposed algorithm does not top the charts for every dataset, it maintains a constant healthy position throughout all the datasets containing diverse, challenging conditions. This performance has been shown by the overall best low RSTE value and largest F-Measure values. The AFTV also performs relatively well, but it requires many computations due to the performance of tedious fusion process for no. of times in a single frame track estimate. This fact is supported by a high time to process details provided in Table 2. Among rest of algorithms, though some perform fast (CPF TBF) but lack in the accuracy counterpart. These provide good results only for a few sequences but perform poorly for other videos. Hence, their overall RSTE and F-Measure in Table 3 and Table 4 are not appreciable respectively.

Table 2. Time to Process (in sec/frame)

Algorithm used	Video Sequence Used				
	Seq. 2	Seq. 4	Seq. 6	Snow Parking	Average
PA	*.25*	*.23*	*.20*	*.30*	*.25*
CPF	**.22**	**.22**	**.15**	**.29**	**.22**
CADF	.59	.56	.43	1.1	.67
AFTV	.90	.80	.65	1.3	.91
TBF	.25	.26	.22	.39	.28

Table 3. The RSTE values

Algorithm used	Video Sequence Used				
	Seq. 2	Seq. 4	Seq. 6	Snow Parking	Average
PA	*.71*	**.78**	**.77**	**.81**	**.77**
CPF	.16	.69	.13	.51	.37
CADF	.44	.70	.14	.19	.37
AFTV	**.77**	.66	**.77**	*.63*	*.71*
TBF	.30	*.72*	*.17*	.61	.45

Table 4. The F-Measure values

Algorithm used	Video Sequence Used				
	Seq. 2	Seq. 4	Seq. 6	Snow Parking	Average
PA	*.71*	**.78**	**.77**	**.81**	**.77**
CPF	.16	.69	.13	.51	.37
CADF	.44	.70	.14	.19	.37
AFTV	**.77**	.66	**.77**	*.63*	*.71*
TBF	.30	*.72*	*.17*	.61	.45

4. Conclusion

In the proposed algorithm, a unique way of tracking by the combined use of visible imaging and thermal imaging was presented. The method formulated a decisive fusion strategy by analyzing the proximity between parameters found from both the imaging modalities. The parameters used for developing the fusion strategy comprised of the average likelihood weight of particles (obtained from individual imagery) and the location value in image space pointed by the track output of individual imaging tracker. It is clear from the experiments that the presented algorithm performed efficiently under various challenging conditions. It provided commendable accuracy along with consuming small processing time as compared with the state-of-the-art methods. The proposed method top scored in case of two sequences and a second position in the rest two data sequences taken during experiments when root square tracking error was found and compared with different state of the art methods. Further the overlapping efficiency in terms of F-Measure was best for three sequences out of four. The proposed method fared with overall RSTE of 4.77 and an F-Measure of 0.77 when averaged over various data sequences. Time consumed was also less in comparison to state of art fusion based trackers and lagged behind only the individual colour image based particle filter. In addition, the fusion scheme also appears to have the advantage over other techniques in term of the fact that it is independent of the tracking algorithm chosen at the individual image sensor level. It requires only two parameters from the output of any tracker: i) Its track confidence measure and, ii) the position of the output track estimate of the tracker. Both of these traits are readily available for an individual tracker. In future, the method can be applied to tracking under bimodal fusion of imageries through any tracker, but more work needs to be done in the direction to formulate scaled versions of such fusion techniques that can be applied more than two modalities.

References

[1] Ali, U., Ali, M.: 'Optimized Visual and Thermal Image Fusion for Efficient Face Recognition' in '9th Int. Conf. on Inf. Fusion', (2006), pp. 1–6.

[2] G. Bebis, A. Gyaourova, S.S. and I.P.: 'Face Recognition by Fusing Thermal Infrared and Visible Imagery' Image Vis. Comput., 2006, 24, (7), pp. 727–742.

[3] Heo, J., Kong, S.G., Abidi, B.R., Abidi, M.A.: 'Fusion of visual and thermal signatures with eyeglass removal for robust face recognition', in 'IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops', USA, (2004).

[4] Kong, S.G., Heo, J., Boughorbel, F., et al.: 'Multiscale fusion of visible and thermal IR images for illumination- invariant face recognition'Int. J. Comput. Vis., 2007, 71, (2), pp. 215–233.

[5] Wilhelm, T., Böhme, H.J., Gross, H.M.: 'A multi-modal system for tracking and analyzing faces on a mobile robot'Rob. Auton. Syst., 2004, 48, (1), pp. 31–40.

[6] Cielniak, G., Duckett, T.: 'Active People Recognition using Thermal and Grey Images on a Mobile Security Robot' IEEE/RSJ Int. Conf. Intell. Robot. Syst. IROS, (2005), pp. 3610–3615.

[7] Cielniak, G., Duckett, T., Lilienthal, A.J.: 'Improved data association and occlusion handling for vision-based people tracking by mobile robots', in '2007 IEEE/RSJ International Conference on Intelligent Robots and Systems' (2007), pp. 3436–3441

[8] Palmerini, G.B., Università, S.: 'Combining Thermal and Visual Imaging in Spacecraft Proximity Operations'2014, 2014, (December), pp. 10–12.

[9] Tong, Y., Liu, L., Zhao, M., Chen, J., Li, H.: 'Adaptive fusion algorithm of heterogeneous sensor networks under different illumination conditions'Signal Processing, 2016, 126, pp. 149–158.

[10] Zhou, Z., Wang, B., Li, S., Dong, M.: 'Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters'Inf. Fusion, 2016, 30, pp. 15–26.

[11] Ma, J., Ma, Y., Li, C.: 'Infrared and visible image fusion methods and applications: A survey'Inf. Fusion, 2018, 45, (December 2017), pp. 153–178.

[12] Walia, G.S., Kapoor, R.: 'Recent advances on multicue object tracking: a survey'Artif. Intell. Rev., 2016, 46, (1), pp. 15821–15847.

[13] Li, C., Wu, X., Zhao, N., Cao, X., Tang, J.: 'Fusing two-stream convolutional neural networks for RGB-T object tracking'Neurocomputing, 2017.
[14] Nummiaro, K., Koller-Meier, E., Van Gool, L.: 'An adaptive color-based particle filter'Image Vis. Comput., 2003, 21, (1), pp. 99–110.
[15] Talha, M., Stolkin, R.: 'Particle filter tracking of camouflaged targets by adaptive fusion of thermal and visible spectra camera data'IEEE Sens. J., 2014, 14, (1), pp. 159–166.
[16] Conaire, C.Ó., O'Connor, N.E., Smeaton, A.: 'Thermo-visual feature fusion for object tracking using multiple spatiogram trackers'Mach. Vis. Appl., 2008, 19, (5–6), pp. 483–494.
[17] Xiao, G., Yun, X., Wu, J.: 'A new tracking approach for visible and infrared sequences based on tracking-before-fusion'Int. J. Dyn. Control, 2016, 4, (1), pp. 40–51.
[18] Xiao, J., Stolkin, R., Oussalah, M., Leonardis, A.: 'Continuously Adaptive Data Fusion and Model Relearning for Particle Filter Tracking With Multiple Features'IEEE Sens. J., 2016, 16, (8), pp. 2639–2649.
[19] Stolkin, R., Rees, D., Talha, M., Florescu, I.: 'Bayesian fusion of thermal and visible spectra camera data for region based tracking with rapid background adaptation'IEEE Int. Conf. Multisens. Fusion Integr. Intell. Syst., 2012, pp. 192–199.
[20] Fendri, E., Boukhriss, R.R., Hammami, M.: 'Fusion of thermal infrared and visible spectra for robust moving object detection'Pattern Anal. Appl., 2017, 20, (4), pp. 907–926.
[21] Singh, S. Khosla, A., and Kapoor, R. 'Visual-Thermal Fusion Based Object Tracking via a Granular Computing Backed Particle Filtering', IETE Journal of Research, 2022, pp. 1-16.
[22] Research, S.R., Chan, A.L.: 'Enhanced target tracking through infrared-visible image fusion'14th Int. Conf. Inf. Fusion, 2011, pp. 1–8.
[23] Singh, S. Khosla, A., and Kapoor, R.,' "Object Tracking with a Novel Visual-Thermal Sensor Fusion Method in Template Matching' International Journal of Image, Graphics and Signal Processing (IJIGSP), 2019, 11(7), pp. 39-47.
[24] Wu, Y., Blasch, E., Chen, G., Bai, L., Ling, H.: 'Multiple source data fusion via sparse representation for robust visual tracking', in '14th International Conference on Information Fusion' (2011), pp. 1–8
[25] Li, C., Cheng, H., Hu, S., Liu, X., Tang, J., Lin, L.: 'Learning Collaborative Sparse Representation for Grayscale-Thermal Tracking'IEEE Trans. Image Process., 2016, 25, (12), pp. 5743–5756.
[26] Li, C., Hu, S., Gao, S., Tang, J.: 'Real-time grayscale-thermal tracking via laplacian sparse representation', in 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)' (2016), pp. 54–65
[27] Liu, H.P., Sun, F.C.: 'Fusion tracking in color and infrared images using joint sparse representation'Sci. China-Information Sci., 2012, 55, (3), pp. 590–599.
[28] Yong, H., Meng, D., Zuo, W., Zhang, L.: 'Robust Online Matrix Factorization for Dynamic Background Subtraction'IEEE Trans. Pattern Anal. Mach. Intell., 2017.

Authors' Profiles

Satbir Singh obtained his Ph.D. degree from the National Institute of Technology, Jalandhar (NITJ), and received M.E. in Electronics and Communication Engineering from Thapar University, Patiala, India. He is presently working with Centre for Artificial Intelligence, NITJ. Previously, he had worked with Central Scientific Instruments Organization - India, Delhi Technological, University, Delhi - India, and Centre of Advanced Computing, Mohali - India. His research interests include computer vision, artificial intelligence, image and signal processing, and IoT.

Arun Khosla received his Ph.D. degree from Indraprastha University, Delhi in Information Technology. He is presently working as Professor in the Department of Electronics, and Communication Engineering, National Institute of Technology, Jalandhar, India. Dr. Khosla has been a reviewer for various IEEE and other National and International conferences and serves on the editorial board of the International Journal of Swarm Intelligence Research. He is a life member of Indian Society of Technical Education.

Rajiv Kapoor received an M.E. and Ph.D. degree in ECE from Delhi College of Engineering, Delhi University, and Punjab University, Chandigarh, respectively. Dr. Kapoor is presently working as Professor in Electronics Communication Engineering Department, Delhi Technological University, Delhi). He has authored over 100 research papers in various renowned international journals and conferences. His primary research interests are machine learning, computer vision, signal and image processing.

How to cite this paper: Satbir Singh, Arun Khosla, Rajiv Kapoor, "Object tracking via a Novel Parametric Decisions based RGB-Thermal Fusion", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.15, No.4, pp. 1-18, 2023. DOI:10.5815/ijigsp.2023.04.01

General Relativity and Gravitation (2023) 55:95
https://doi.org/10.1007/s10714-023-03140-5

RESEARCH

Observational constraints on teleparallel effective equation of state

S. H. Shekh[1] · H. Chaudhary[2] · A. Bouali[3] · A. Dixit[4]

Received: 15 April 2023 / Accepted: 4 August 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

A cluster of astronomical observations emerged that the present Universe is in a phase of accelerated expansion. To examine the accelerated expansion of the Universe, we used the newly proposed gravity in which torsion scalar is responcible for the accelerated expansion. The investigation is carried out using a parameterized effective equation of state with the free parameters α, β, ω_0 and ω_1. Also, for the analysis we consider the model of the form $f(T) = m(-T)^n$, where $m \neq 0$ and n are the constants. By constraining the model with the combined $H(z)$, SC, $unCorBAO$ sample datasets. The model appeared to be in good agreement with the observations, and we were able to determine the parameter values that fit the data the best. Finally, we examined how the deceleration parameter and equation of state parameter behaved in relation to the parameter values that best suited the data.

Keywords FRW universe · Effective equation of state · $f(T)$ gravity · Cosmology

✉ S. H. Shekh
da_salim@rediff.com

H. Chaudhary
himanshuch1729@gmail.com

A. Bouali
a1.bouali@ump.ac.ma

A. Dixit
archana.dixit@gla.ac.in

1 Department of Mathematics, S.P.M. Science and Gilani Arts, Commerce College, Ghatanji, Yavatmal, Maharashtra 445301, India

2 Department of Applied Mathematics, Delhi Technological University, New Delhi, Delhi 110042, India

3 Laboratory of Physics of Matter and Radiation, Mohammed I University, BP 717, Oujda, Morocco

4 Department of Mathematics, GLA University, Mathura, U.P. 281406, India

Published online: 23 August 2023

Contents

1 Introduction
2 $f(T)$-gravity
3 Components of field equations with isotropic background
4 Cosmological model and equation of state
4.1 Deceleration parameter
4.2 Statefinder diagnostic
4.3 $O_m(z)$ diagnostic
4.4 Energy conditions
5 Data description
5.1 Hubble datasets
5.2 Standard candles (SC)
5.3 Uncorrelated baryon acoustic oscillations (unCor BAO)
6 Model-I
6.1 Cosmological parameters
6.1.1 The equation of state parameter
6.1.2 The deceleration parameter
6.1.3 Statefinder diagnostics
6.1.4 $O_m(z)$ diagnostic
6.1.5 Energy conditions
7 Model-2
7.1 Cosmological parameters
7.1.1 The equation of state parameter
7.1.2 The deceleration parameter
7.1.3 Statefinder diagnostics
7.1.4 $O_m(z)$ diagnostic
7.1.5 Energy conditions
8 Information criteria
9 Age of universe
10 Conclusions
References

1 Introduction

After astronomical observations of supernovae over the past few decades [1, 2], multiple ideas to change the General Relativity (GR) and to take into account various formulations of gravity have been made. There are several different modified theories of gravity are present, including $f(R)$, which is produced by substituting an arbitrary function of the Ricci scalar for the Ricci scalar [3–5], $f(R, T)$ gravity is obtained by changing the Ricci scalar and the trace of the energy tensor [6–10] with any arbitrary function, The Gauss–Bonnet invariant has a generic function $f(G)$ [11, 12]. $f(R, G)$ which combines Ricci scalar and Gauss–Bonnet scalar [13, 14], $f(T, B)$ which involve torsion and boundary term [15]. Next, the Teleparallel Gravity (TG) is one of the modification of Einstein's theory that uses the Weitzenbock connection rather than the Levi–Civita connection. As a result, it has no curvature but has torsion, which is what causes the acceleration of the universe. Torsion is responsible for the universe's accelerated expansion and is made up of four parallel vector fields known as vierbeins that are linearly independent. A significant advantage about this theory is that its field equations are second order, which makes them easier to work with than those in $f(R)$ theory. In addition, Wu and Yu [16] point out that the field

equations of $f(T)$ theory are second order, making them much easier to solve than the corresponding fourth-order equations of $f(T)$ theory. The teleparallel theory of gravity, which is equivalent to general relativity, is the fundamental basis of the $f(T)$ theory [17, 18]. $f(T)$ modified teleparallel gravity has been researched by Karami and Abdolmaleki [19] as a substitute for holographic and new-agegographic dark energy models. Amir and Yussouf developed the Kantowski-Sachs universe models in the $f(T)$ theory of gravity [20].To examine the cosmological evolution's of the equation of state parameter for dark energy and their observational constraints, Bamba et al. [21] recently discussed various $f(T)$ models. Chen et al. [22] examined this issue by extending modified gravity at both the background and perturbed levels also examined this theory in terms of the quintessential scenario.

Observational data indicate that the expansion of the universe has accelerated. Einstein's general relativity proposes the concept of Dark Energy (DE) to explain this accelerated expansion of the universe. DE is a fluid with negative pressure and positive energy density, so its equation of state (EoS) is negative: $\omega = \frac{p}{\rho}$. There are many ways to explain the nature of various dark energy models. Authors [23, 24] have provided a presentation for ω that was observed from experimental data. The EoS parameter is predicted to be $\omega = -1$ by DE models with a cosmological constant (ΛCDM), whereas quintessence models predict $\omega > -1$. However, recent observational data favour phantom models with $\omega < -1$ over quintessence models [25]. The 9 year WMAP survey suggests the EoS parameter to be "$\omega = -1.073^{+0.090}_{-0.089}$ from CMB measurements and $\omega = -1.084 \pm 0.063$ in combinations with Supernova data [8]. Kumar and Xu from a combined analysis of the data sets of SNLS3, BAO, Planck, WMAP9 and WiggleZ constrained the EoS parameter as $\omega = -1.06^{+0.11}_{-0.13}$ [26]. Amanullah et al. of Supernova cosmology project have found that $\omega = 1.1035^{+0.055}_{-0.058}$ [27]. Moreover the recent Planck 2018 results constrained $\omega = -1.03 \pm 0.03$" [28]. In light of these constraints on the EoS parameter derived from observations, a phantom phase in the universe cannot be ruled out. In phantom model, energy conditions are violated and the universe evolves to a future singularity in a finite amount of time. According to the classification of Nojiri et al. [29], there are four distinct singularity possibilities [30]: (i) Sudden singularity (type I singularity), in which the pressure becomes infinite,(ii) Big Rip singularity (type-II singularity) where the scale factor and density becomes infinite in finite time Type-III singularity, in which both the pressure and density become infinite while the scale factor remains finite, and (iii) Type-IV singularity, in which the higher derivatives of the Hubble parameter diverge [31].

If we look at the universe's expansion history, one can see that some cosmological parameters play an important role in designating the cosmological evolution. It is well known that the equation of state parameter ω predicts various fluid descriptions of the universe. Hubble $H(z)$ dimensions are also used to derived some traditional cosmological parameters, like "the density of dark energy and the density of non-relativistic matter" [32–34]. These provides constraints which are good in comparison with SN-Ia data, but not are good as provided by BAO and CMB observations. $H(z)$ data has recently been utilized to calculate the Hubble constant H_0 [35, 36], the derived value of H_0, being more compatible with recent lower values measured from a

"median statistics analysis of Huchra's H_0 compilation". All of the above findings were discovered by a simple analysis of "57 supernova data points". Many Authors [37–39] compiled updated list of 38 measurements of Hubble parameters $H(z)$ between redshifts $0 \leq z \leq 2.36$ and used them to put constraints on model parameters. Authors [40] used $H(z)$ data, pantheon compilation of SN-Ia data and BAO data to constrained the model parameters using χ^2 minimization technique. They also obtained the present values of Hubble constant H_0 as $66.2^{+1.42}_{-1.34}$,$70.7^{+0.32}_{-0.31}$ and $67.74^{1.24}_{-1.04}$ for $H(z)$, $H(z)+Pantheon$ and $H(z)+BAO$ respectively [40, 41]. The pantheon compilation of SN-Ia data in the redshift range $0.01 < z < 2.3$ is given in Scolnic et al. [42]

In this model we have use the MCMC methods for quantitative analysis. As a result of the discussion and observations made above, we were motivated to study the cosmological models with effective equation of state parameter in $f(T)$ theory of gravit under "observational constraints of effective equation of state with $f(T)$ gravity".

2 $f(T)$-gravity

We provide a succinct explanation of $f(T)$ gravity in this section, along with a thorough derivation of its field equations. Let's define the Greek and Latin notations of the Latin subscript as those connected to the space-time coordinates and the tetrad field, respectively. We can define the line element for a general space-time metric as

$$dS^2 = g_{\mu\nu}dx^\mu dx^\nu, \tag{1}$$

where the metric's ten degrees of freedom and symmetry are represented by the variables $g_{\mu\nu}$. The theory can be expressed either in space-time or in tangent space, allowing us to recast the line element that can be transformed into the tetrad described by Minkowski (which represents the dynamic fields of the theory) as follows [16]

$$dS^2 = g_{\mu\nu}dx^\mu dx^\nu = \eta_{ij}\theta^i\theta^j, \tag{2}$$

$$dx^\mu = e_i^\mu \theta^i, \quad \theta^i = e_\mu^i dx^\mu, \tag{3}$$

where η_{ij} is a metric on Minkowski space-time and $\eta_{ij} = diag[1, -1, -1, -1]$ and $e_i^\mu e_\nu^i = \delta_\nu^\mu$ or $e_i^\mu e_\mu^j = \delta_i^j$. The root of metric determinant is given by $\sqrt{-g} = \det[e_\mu^i] = e$. Towards a manifold where the Riemann tensor part without the torsion terms is null (contribution of the Levi–Civita connection) and only the non-zero torsion terms exist, the Weitzenbocks connection components are defined as follows [17]

$$\Gamma^\alpha{}_{\mu\nu} = e_i^\alpha \partial_\nu e_\mu^i = -e_\mu^i \partial_\nu e_i^\alpha. \tag{4}$$

It possesses a nonzero torsion but a zero curvature. From this relationship, the primary geometrical objects of space-time are built. The antisymmetric portion of this connection defines the elements of the tensor torsion through the link as

$$T^{\alpha}{}_{\mu\nu} = \Gamma^{\alpha}{}_{\mu\nu} - \Gamma^{\alpha}{}_{\nu\mu} = e_i^{\alpha}\left(\partial_\mu e_\nu^i - \partial_\mu e_\mu^i\right), \tag{5}$$

The contortion tensor which is a space-time tensor that represents the difference between the Levi–Civita and Weitzenbock connections as:

$$K^{\mu\nu}{}_{\alpha} = \left(-\frac{1}{2}\right)\left(T^{\mu\nu}{}_{\alpha} + T^{\nu\mu}{}_{\alpha} - T^{\mu\nu}{}_{\alpha}\right). \tag{6}$$

In order to make more clear the definition of the scalar equivalent to the curvature scalar of RG, we first define a new tensor $S^{\mu\nu}{}_{\alpha}$, constructed from the components of the tensors torsion and contorsion tensor as

$$S^{\mu\nu}{}_{\alpha} = \left(\frac{1}{2}\right)\left(K^{\mu\nu}{}_{\alpha} + \delta_\alpha^\mu T^{\beta\nu}{}_{\beta} - \delta_\alpha^\nu T^{\beta\mu}{}_{\beta}\right). \tag{7}$$

The torsion scalar T is

$$T = T^{\alpha}{}_{\mu\nu}\, S^{\mu\nu}{}_{\alpha}. \tag{8}$$

Now, we define the action by generalizing the TG i.e. $f(T)$ theory as [18, 19]

$$S = \int \left[f(T) + L_{matter}\right]\, e\, d^4x. \tag{9}$$

Here, $f(T)$ denotes an algebraic function of the torsion scalar T. Making the functional variation of the action (9) with respect to the tetrads, we get the following equations of motion [20]

$$S_\mu^{\nu\rho}\partial_\rho T f_{TT} + \left[e^{-1}e_\mu^i\partial_\rho\left(e e_i^\alpha S_\alpha^{\nu\rho}\right) + T^{\alpha}{}_{\lambda\mu}S_\alpha^{\nu\lambda}\right] f_T + \frac{1}{4}\delta_\mu^\nu\,(f) = 4\pi T_\mu^\nu. \tag{10}$$

The field equation (10) looks considerably different from Einstein's equation since it is expressed in terms of tetrads and partial derivatives.
where T_μ^ν is the energy momentum tensor, $f_T = df(T)/dT$, $f_{TT} = d^2 f(T)/dT^2$. As in our investigations, the energy momentum tensor is defined as

$$T_\mu^\nu = (\rho + p)\, u^\nu u_\mu - p\, g_\mu^\nu, \tag{11}$$

where u^ν is the four velocity vector, the pressure p for this fluid is taken to be isotropic, ρ be the energy density. The energy momentum tensor can be expressed as

$$\begin{aligned} T_\mu^\nu &= diag[-\rho,\ \ p,\ \ p,\ \ p\ ] \\ &= diag[-1,\ \omega,\ \omega,\ \omega\ ]\,\rho, \end{aligned} \tag{12}$$

where ω is the equation of state parameter of the form $\omega = \frac{p}{\rho}$.

3 Components of field equations with isotropic background

To describe models of FRW space-time, let's first establish the equations of motion of a collection of diagonal tetrads using the Cartesian coordinate metric, as

$$ds^2 = -dt^2 + a^2(dx^2 + dy^2 + dz^2), \tag{13}$$

where a be the cosmological scale factor. For this frame the langrangian density or Torsion scalar observed as

$$T = -6\left(\frac{\dot{a}}{a}\right)^2 = -6H^2 \tag{14}$$

where $H = \frac{\dot{a}}{a}$ be the Hubble's parameter.

$$12H^2 f_T(T) + f(T) = 16\pi G\rho \tag{15}$$
$$48H^2\dot{H} f_{TT}(T) - 4(\dot{H} + 3H^2) f_T(T) - f(T) = 16\pi G p \tag{16}$$

where the overhead dot $(\cdot)$ denotes the derivative with respect to cosmic time t, p and ρ denotes the pressure and energy density respectively. The conservation equation leads as

$$\dot{\rho} + 3H(\rho + p) = 0. \tag{17}$$

However, Eqs. (15) and (16) can be rewritten as

$$H^2 = \frac{8\pi G}{3}(\rho + 3\rho_\Lambda) = 0. \tag{18}$$

and

$$2\dot{H} + 3H^2 = -8\pi G(p + 3p_\Lambda). \tag{19}$$

where p_Λ and ρ_Λ are the effective pressure and energy density of the fluid. Specifically, p_Λ and ρ_Λ express in T and f_T are as follows

$$\rho_\Lambda = \frac{1}{16\pi G}[2T f_T(T) - f(T) - T] \tag{20}$$

and

$$p_\Lambda = \frac{1}{16\pi G}[4\dot{H}(2T f_{TT}(T) - f_T(T) - 1)] - \rho_\Lambda \tag{21}$$

4 Cosmological model and equation of state

After solving Eqs. (20) and (21) for p_Λ and ρ_Λ, we obtained the effective or total EoS parameter ($\omega_\Lambda = \frac{p_\Lambda}{\rho_\Lambda}$) as

$$\omega_\Lambda = \frac{p_\Lambda}{\rho_\Lambda} = -1 + \frac{4\dot{H}(2Tf_{TT}(T) - f_T(T) - 1)}{2Tf_T(T) - f(T) - T} \tag{22}$$

For our analysis, we consider the simplest functional form of $f(T)$ gravity model of the form

$$f(T) = m(-T)^n, \tag{23}$$

where $m \neq 0$ and n are the constants. We discovered the total or effective equation of state parameter, as

$$\omega_\Lambda = -1 + \frac{(\alpha - 1)}{3}\frac{\dot{H}}{H^2} \tag{24}$$

Moreover, we may define the relationship for t and z as follows by utilising the relation $\frac{a}{a_0} = \frac{1}{(1+z)}$,

$$\frac{d}{dt} = \frac{dz}{dt}\frac{d}{dz} = -(1+z)H(z)\frac{d}{dz}, \tag{25}$$

By normalizing the present value of the scale factor $a_0 = a(0) = 1$, the Hubble parameter reduces to

$$\dot{H} = -(1+z)H(z)\frac{d}{dz}, \tag{26}$$

One additional alternative equation is required in order to solve Eq. (24) for $H(z)$. Thus, we suppose that the equation of state parameter has a well-motivated parametric form as a function of redshift z as [43, 44].

$$\omega_\Lambda = \frac{-1}{1 + \alpha(1+z)^\beta} \tag{27}$$

and

$$\omega_\Lambda = \omega_0 + \omega_1\left(\frac{z}{1+z}\right) \tag{28}$$

where α, β, ω_0 and ω_1 are the free model parameters.

The above effective equation of states of the universe involves some unknown dimension less constant parameters namely α, β, ω_0 and ω_1. We are going to determine these unknowns from numerical analysis adopted here for different observed data.

Moreover, the observed dimensionless Hubble parameter and a suitable chi-square function are defined in the following sections with different cases. Before that, in regard to the Universe Hubble parameter, some cosmographic/cosmological parameters that are important in cosmological observations are such as

4.1 Deceleration parameter

The deceleration parameter, denoted as $q(z)$, is defined as a function of the Hubble parameter H by the equation:

$$q(z) = -1 - \frac{\dot{H}}{H^2} \tag{29}$$

The deceleration parameter is a key quantity for characterizing the different phases of the universe. It helps distinguish between decelerated phases ($q > 0$) and accelerated phases ($q < 0$). Understanding the evolution of the universe requires insights into the behavior of the deceleration parameter. In the field of cosmology, various studies have explored the significance of the deceleration parameter in elucidating the different phases of the universe. For instance, references such as [45] have employed the deceleration parameter to analyze and explain the evolutionary phases of the universe in different cosmological models. Moreover, researchers have also worked on parameterizing the deceleration parameter to study its behavior in various scenarios [46–51]. These parameterizations allow for a more comprehensive understanding of the underlying dynamics and potential transitions between decelerated and accelerated phases.

4.2 Statefinder diagnostic

In the analysis of cosmology, Sahni et al. [52] have included a new cosmological diagnostic pair $\{r, s\}$ called Statefinder diagnostic. The Statefinder help to differentiate and compare between different dark energy models. The standard cold dark model (SCDM) and the cosmological constant model (ΛCDM) have some fixed points in the $\{r - s\}$ plane and $\{r - q\}$ plane. The standerd equations of r and s are

$$r = q(z) + 2q^2(z) - \frac{\dot{q}}{H} \quad \text{and} \quad s = \frac{(r-1)}{3(q-1/2)} \tag{30}$$

4.3 $O_m(z)$ diagnostic

Now, we going towards the discussion on $O_m(z)$ diagnostic which is used to differentiate the standard ΛCDM model from various dark energy dominated models [53]. In accordance with [54, 55] for flat universe $O_m(z)$ is defined as

$$O_m(z) = \frac{\left(\frac{H(z)}{H_0}\right)^2 - 1}{(1+z)^3 - 1} \tag{31}$$

4.4 Energy conditions

Without specifying the specific behaviour of the matter, the energy conditions allow us to consider the behaviour of gravitational systems. It has been demonstrated that the well-known Raychaudhuri equation, which deals with interesting gravity, is very useful in characterising energy circumstances. The four point-wise energy conditions that are limitations on the stress-energy-momentum tensor of the matter are strong energy condition (SEC), weak energy condition (WEC), null energy condition (NEC), and dominant energy condition (DEC). In physical these energy conditions are defined as

- "Null energy condition (NEC) $\iff \rho + p \geq 0$,
- Weak energy condition (WEC) $\iff \rho + p \geq 0$ and $\rho \geq 0$,
- Strong energy condition (SEC) $\iff \rho + p \geq 0$ and $\rho + 3p \geq 0$,
- Dominant energy condition (DEC) $\iff \rho - |p| \geq 0$ and $\rho \geq 0$".

It is well recognized that among the energy conditions, WEC is the most unpredictable. The WEC assumes that for a perfect fluid, $\rho \geq 0$. Moreover, the universe's pressure cannot be so low that it overwhelms the energy density, or $\rho + p \geq 0$. Next, the null energy condition yields the second law of black hole thermodynamics (NEC). According to the context, we may also deduce that the WEC and DEC, on the other hand, cannot be satisfied if the NEC is violated. In several modified theories of gravity, the energy conditions are investigated, including Capozziello [56] and Alvarenga [57] used the power law in $f(R)$ and $f(R, T)$ gravity to study the phenomenon. In his research, Liu [58] used both Born-Infeld and exponential $f(\mathcal{T})$ gravity, in $f(\mathcal{G})$ gravity by Garcia [59] and Bamba [60] also Atazadeh [61] investigated in $f(R, \mathcal{G})$ gravity while Mishra et al. has looked at the energy conditions in cosmological models with changing anisotropic parameters in $f(R, T)$ gravity. [62].

5 Data description

5.1 Hubble datasets

It is possible to write the Hubble parameter as $\frac{d}{dz} = \frac{d}{dz}\frac{dz}{dt} = -(1+z)H(z)\frac{d}{dz}$. The model-independent value of the Hubble parameter can be determined by measuring dt since dz is derived from a spectroscopic survey. Here, we make use of the $H(z)$ data point collection, which consists of 57 points (see Table I in [63]). measured using the differential age technique [64]. The chi-square function for $H(z)$ defined as

$$\chi^2_{H(z)} = \sum_{i=1}^{57} \frac{[H_{th}(z_i) - H_{obs}(z_i)]^2}{\sigma^2_{H(z_i)}}. \tag{32}$$

Here, the Hubble parameter's theoretical value is represented by H_th, its observed value is represented by H_obs, and the standard error of the observed value of H is represented by $\sigma_{H(z_i)}$.

5.2 Standard candles (SC)

Considering the measurements of the Pantheon Type Ia supernova dataset [65–69], Quasars [70] and Gamma Ray Bursts were collected in [71]. The model parameters can be fitted by comparing the observed μ_i^{obs} with theoretical μ_i^{th} value of the distance moduli is given by

$$\mu = m - M = 5\log_{10}(D_L) + \mu_0, \tag{33}$$

where M and m represents the absolute and apparent magnitudes respectively and $\mu_0 = 5\log\left(H_0^{-1}/Mpc\right) + 25$ is the nuisance parameter that has been depreciate. The luminosity distance is defined by

$$D_L(z) = \frac{c}{H_0}(1+z)\int_0^z \frac{dz^*}{E(z^*)} \tag{34}$$

The χ^2 function of the (SNeIa) measurements is given as

$$\chi^2_{\text{SN}}\left(\phi_{\text{s}}^{\nu}\right) = \mu_{\text{s}}\mathbf{C}^{-1}_{\text{s, cov}}\,\mu_{\text{s}}^T, \tag{35}$$

where $\mu_{\text{s}} = \{\mu_1 - \mu_{\text{th}}(z_1, \phi^{\nu}), \ldots, \mu_N - \mu_{\text{th}}(z_N, \phi^{\nu})\}$ and the subscript 's' denotes SnIa. Keep in mind for SnIa dataset the covariance matrix is not diagonal. The distance modulus μ_i is given as $\mu_i = \mu_{B,i} - \mathcal{M}$, where $\mu_{B,i}$ represents the maximum apparent magnitude in the cosmic rest frame for redshift z_i and $\mathcal{M}$ represents the universal free parameter, by various observational ambiguity both $\mathcal{M}$ and h parameters are naturally degraded with respect to Pantheon dataset, Hence one can't extract proper information about H_0 from SnIa data alone.

5.3 Uncorrelated baryon acoustic oscillations (unCor BAO)

Considering the largest collection of BAO dataset measurements, from it we have selected 17 uncorrelated BAO measurements because using the complete collection of BAO might lead to a very large error due to data correlations, therefore we select the small dataset to reduce errors. Studies of BAO in transverse direction contribute a measurement of $D_H(z)/r_d = c/H(z)r_d$ with comoving angular diameter distance [72].

$$D_M = \frac{c}{H_0} S_k\left(\int_0^z \frac{dz'}{E(z')}\right), \tag{36}$$

where

$$S_k(x) = \begin{cases} \frac{1}{\sqrt{\Omega_k}}\sinh\left(\sqrt{\Omega_k}x\right) & \text{if} \quad \Omega_k > 0 \\ x & \text{if} \quad \Omega_k = 0 \\ \frac{1}{\sqrt{-\Omega_k}}\sin\left(\sqrt{-\Omega_k}x\right) & \text{if} \quad \Omega_k < 0. \end{cases} \tag{37}$$

Table 1 Summary of the MCMC results using $H(z) + SC + unCorBAO$ dataset

MCMC results Model	Parameters	Best fit value
ΛCDM Model	H_0	$69.854848^{+1.259100}_{-1.259100}$
Model I	H_0	$68.990853^{+0.502428}_{-0.502428}$
	α	$0.297626^{+0.030564}_{-0.030564}$
	k	$-0.179266^{+0.014411}_{-0.014411}$
	β	$6.713197^{+0.474649}_{-0.474649}$

We also take into account the $D_A = D_M/(1+z)$ and the $D_V(z)/r_d$ angular diameter distances. This combines the coordinates of the BAO peak with the sound horizon at the drag epoch, r_d. Ultimately, we are able to directly observe the Hubble parameter using "line-of-sight" (or "radial" observations), To see more please see [73]

$$D_V(z) \equiv \left[z D_H(z) D_M^2(z) \right]^{1/3}. \tag{38}$$

6 Model-I

To solve Eq. (24) for $H(z)$, we assume a well-motivated parametric form of equation of state parameter as a function of redshift z. Using Eqs. (24), (25) and (26), we obtained the Hubble parameter $H(z)$ in terms of redshift z as,

$$H(z) = \frac{H_0}{[1+\alpha(1+z)^{\beta}]^k} \tag{39}$$

where $k = \frac{3}{n(1-m)}$.
Here, we obtained the constraints values of the constants involve in Eq. (39) by combining the above three chi-square functions defined in Eqs. (32), (33) and (34). We have discovered the parameters best fit predicted values as $H_0 = 69.854848^{+1.259100}_{-1.259100}$, $\alpha = 0.297626^{+0.030564}_{-0.030564}$, $\beta = 6.71319^{+0.474649}_{-0.474649}$ and $k = 0.17926^{+0.01441}_{-0.01441}$ shown in Table 1 where as the theoretical curve of Hubble function and distance modulus with ΛCDM model with their corresponding error bars is most likely seen in Figs. 1 and 2 and the likelihood contours for the same parameters at 1σ, 2σ level is shown in Fig. 3.

6.1 Cosmological parameters

6.1.1 The equation of state parameter

Figure 4, shows the EoS parameter's graphical behaviour quite clearly. The equation of state parameter in our model exhibits a transition from a positive to a negative value with cosmic evolution, indicating that the variable equation of state parameter is what causes the change from radiation-dominated matter to $-1 < \omega_\Lambda$ and $-1 > \omega_\Lambda$.

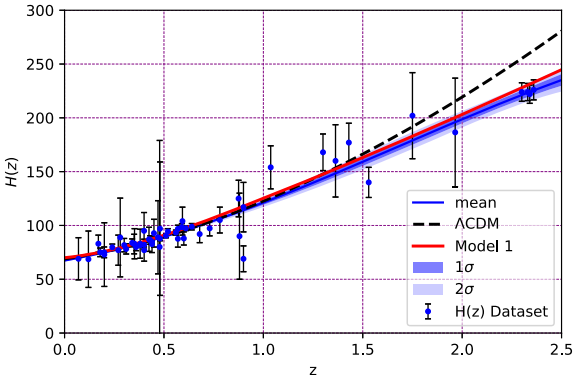


Fig. 1 The figure shows the theoretical curve of Hubble function $H(z)$ of Model 1 and ΛCDM models, shown in blue dots with their corresponding error bars

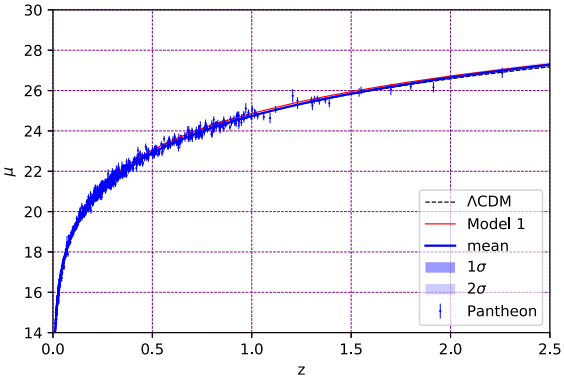


Fig. 2 The figure shows the theoretical curve of distance modulus $\mu(z)$ of our model and ΛCDM models against the Pantheon dataset, shown in pink dots with their corresponding error bars

The model begins in a matter-dominated era, fluctuates in the quintessence region $-1 < \omega_\Lambda < \frac{1}{-3}$, and finally approaches the phantom region $\omega_\Lambda < -1$. The universe is currently dominated by ΛCDM, as indicated by the current value of the equation of state parameter $\omega_\Lambda = -1$. It has been noted that at the current time $z = 0$, it is located in the quintessence region $(-1 \geq \omega_\Lambda)$. Also, the fact that the current values of the EoS parameter, $\omega_0 = -0.6302$ for the $H(z) + SC + unCorBAO$ dataset, are in good accord with recent Planck observational data is another surprising outcome of our model along with Aghanim et al. provide the restrictions on the EoS value for DE in [75]. It is clear that our predictions of the EoS parameter's current value fall within the observational bounds indicated in the previous reference. It is important to note that our findings for the future are identical to the observations made in [76].

Also, one observed that in Fig. 4, the initial value of ω_Λ as $(z \longrightarrow \infty)$ is 0.76547. The default value of ω as $(z \longrightarrow 0)$ is $\omega_\Lambda = -0.6302$, and the far future value is $(z \longrightarrow -1)$ is -1.

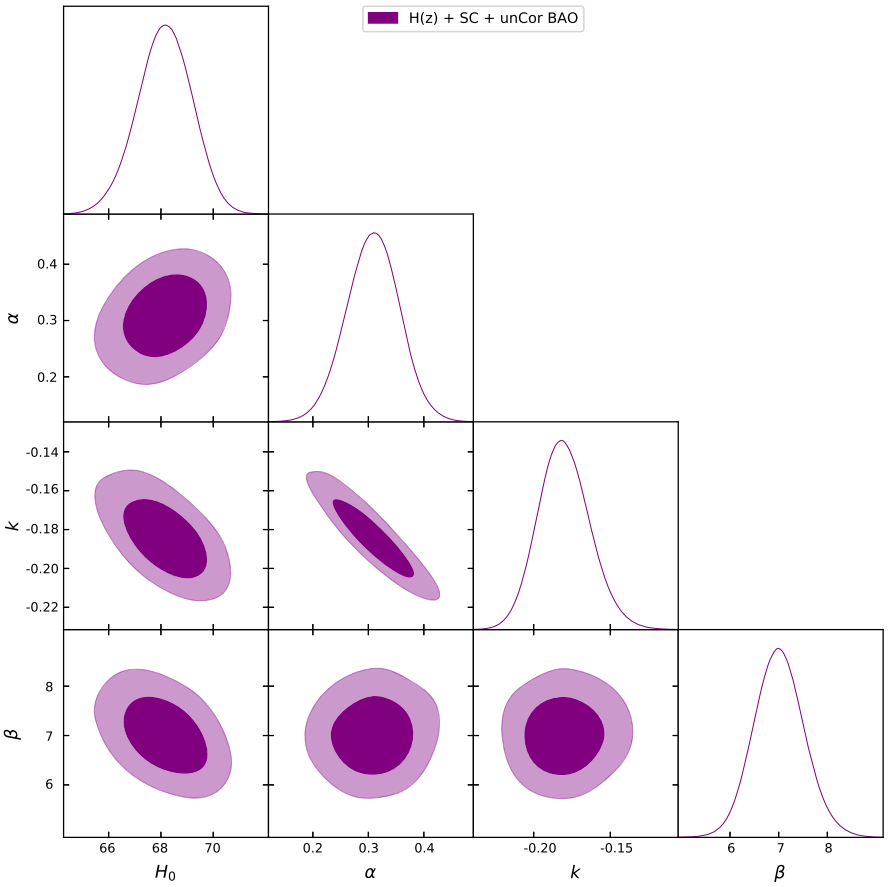


Fig. 3 This figure corresponds to 1σ and 2σ confidence contours obtained from $H(z)+SC+unCorBAO$ dataset obtained for the emergent model

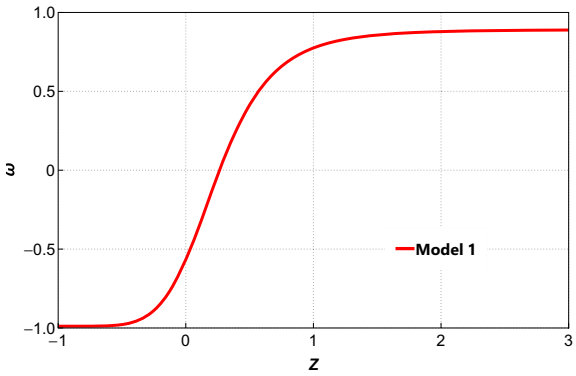


Fig. 4 Evolution of EoS parameter of Model 1 with respect to redshift

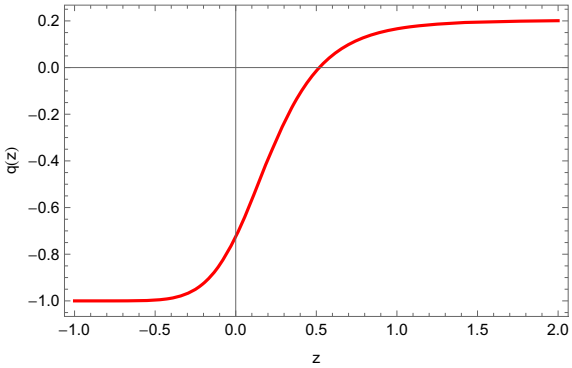


Fig. 5 Evolution of deceleration parameter of Model 1 with respect to redshift

6.1.2 The deceleration parameter

Using Eq. (39) in Eq. (29), the deceleration parameter observed as

$$q(z) = \frac{\alpha\beta k(-z-1)(z+1)^{\beta-1}}{\alpha(z+1)^{\beta}+1} - 1 \tag{40}$$

The behavior of redshift dependendant deceleration parameter $q(z)$ is shown in Fig. 5. This parameter initial setting (q_i as $z \longrightarrow \infty$) is $q = 0.2$, the default deceleration parameter value (q_0 as $z \longrightarrow 0$) is $q = -0.75013$ and thus the deceleration parameter's distant future value (q_{f} as $z \longrightarrow -1$) is $q = -1$. The redshift during which the Universe transitioned from decelerating to accelerating expansion, for this model z_t for which $q = 0$, is 0.5013. The model seems to have a super-accelerated evolution. With a transition redshift of $z_t = 0.513$, which is comparable to the transition redshift of $z_t = 0.50 - -0.73$ observed for the ΛCDM model [74, 75], the cosmos smoothly transitioned from the decelerated phase to the accelerating phase. Also, The current behaviour perfectly corresponds to the analysis of [76] probe using an anisotropic model and a decaying cosmological term for $n = 1.75$.

6.1.3 Statefinder diagnostics

Using Eq. (39) in Eq. (30), the statefinder diagnostics parameters observed as

$$r = \frac{1}{\left(\alpha(z+1)^{\beta}+1\right)^2}\alpha(z+1)^{\beta} \times (-(\beta-3)\beta k + \alpha(\beta k+1)(2\beta k+1)(z+1)^{\beta}+2)+1 \tag{41}$$

$$s = \frac{1}{3}\beta\left(\frac{2k+1}{\alpha(z+1)^{\beta}+1} - \frac{3}{\alpha(2\beta k+3)(z+1)^{\beta}+3} - 2k\right) \tag{42}$$

The behavior of $\{r, s\}$ and $\{r, q\}$-planes is well described in Figs. 6 and 7 respectively. At present the $\{r, s\}$-plane well defines the ΛCDM model with $s = 0$ and $r = 1$ (see Fig. 6) and advances to the quintessence domain before rebounding to the Chaplygin

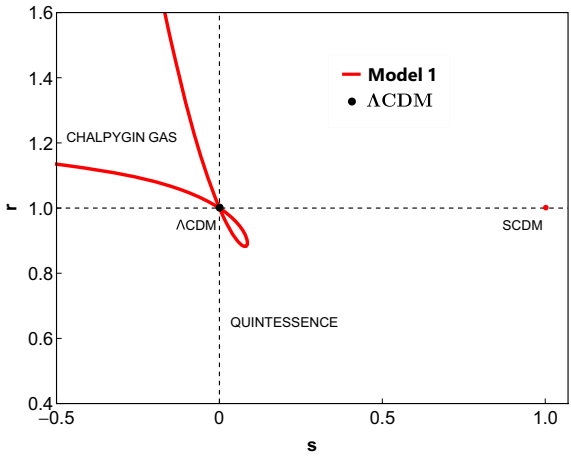


Fig. 6 Behavior of $\{r, s\}$-plane with respect to redshift

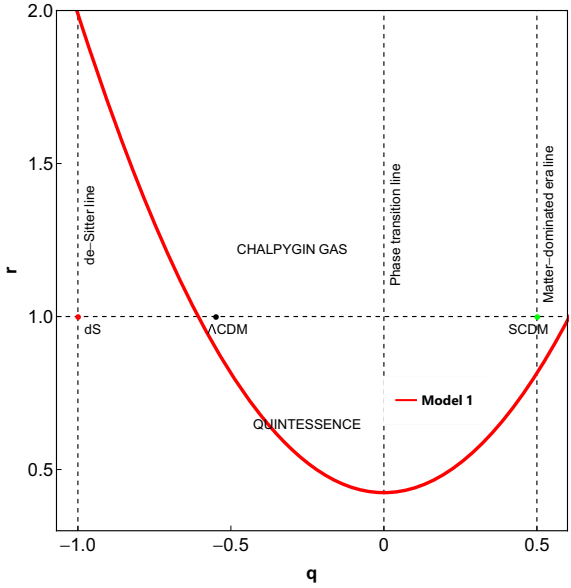


Fig. 7 Behavior of $\{r, q\}$-plane with respect to redshift

gas domain at late times by crossing the intermediate ΛCDM fixed point {0, 1} during expansion while the $\{r, q\}$-plane provides extra information regarding of this model that it evolves from quintessence region in the past and goes to Chaplygin gas (see Fig. 7)

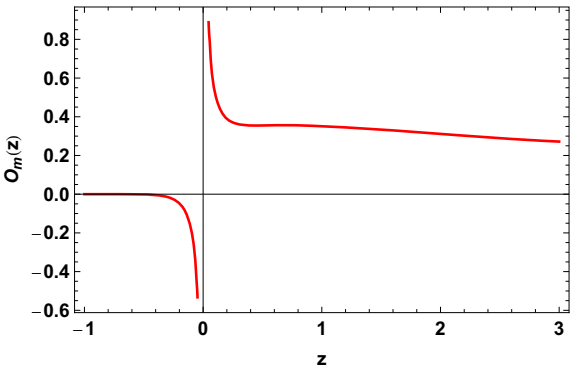


Fig. 8 The evolution of the $O_m(z)$ diagnostic parameter with respect to redshift

6.1.4 $O_m(z)$ diagnostic

Using Eq. (39) in Eq. (31), the $O_m(z)$ diagnostic parameter observed as

$$O_m(z) = \frac{\left(\alpha(z+1)^{\beta}+1\right)^{-2k}-1}{(z+1)^3-1} \tag{43}$$

The behavior of $O_m(z)$ diagnostic parameter verses redshift z is observed in Fig. 8. The $O_m(z)$ value is very small in the redshift range of $z > 0$, indicating that the models are in the quintessence domain. As z declines, the value drops and becomes negative, signifying that the model inhabit the phantom region (see Fig. 8).

6.1.5 Energy conditions

Using Eq. (39), the expressions of NEC, DEC and Strong energy condition respectively observed as,

$$\rho_\Lambda + p_\Lambda = -\frac{2\alpha\beta k(-z-1)(z+1)^{\beta-1}\left(-m6^n n(2n-3)\left(H_0^2\left(\alpha(z+1)^{\beta}+1\right)^{-2k}\right)^n - 6H_0^2\left(\alpha(z+1)^{\beta}+1\right)^{-2k}\right)}{3\left(\alpha(z+1)^{\beta}+1\right)} \tag{44}$$

$$\begin{aligned}\rho_\Lambda - p_\Lambda &= \frac{2}{3}\left(\alpha(z+1)^{\beta}+1\right)^{-2k-1}\left(H_0^2 m6^n\left(\alpha(z+1)^{\beta}(n(\beta k(2n-3)+6)-3)+6n-3\right)\right.\\ &\quad\left.\times\left(H_0^2\left(\alpha(z+1)^{\beta}+1\right)^{-2k}\right)^{n-1}+6H_0^2\left(\alpha(\beta k+3)(z+1)^{\beta}+3\right)\right)\end{aligned} \tag{45}$$

$$\begin{aligned}\rho_\Lambda + 3p_\Lambda &= -2\left(\alpha(z+1)^{\beta}+1\right)^{-2k-1}\left(H_0^2 m6^n\left(\alpha(z+1)^{\beta}(n(\beta k(2n-3)+2)-1)+2n-1\right)\right.\\ &\quad\left.\times\left(H_0^2\left(\alpha(z+1)^{\beta}+1\right)^{-2k}\right)^{n-1}+6H_0^2\left(\alpha(\beta k+1)(z+1)^{\beta}+1\right)\right)\end{aligned} \tag{46}$$

In the following Fig. 9, the behaviour of the current cosmological model's Null (red curve), Dominant (black curve), and Strong (green curve) energy conditions are depicted. The figure explains that the NEC ($\rho_\Lambda + p_\Lambda$) and DEC ($\rho_\Lambda - p_\Lambda$) both energy conditions are verified while SEC ($\rho_\Lambda + 3p_\Lambda$) is violated.

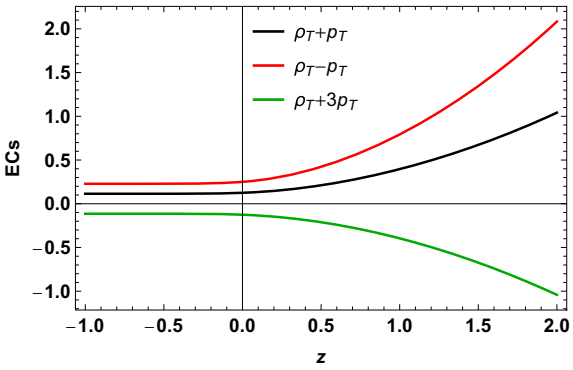


Fig. 9 Evolution of energy conditions with respect to redshift

Table 2 Summary of the MCMC results using $H(z)+SC+unCorBAO$ dataset

MCMC results		
Model II	H_0	$62.984292^{+0.415798}_{-0.415798}$
	β	$0.918487^{+0.009755}_{-0.009755}$
	β_0	$1.018775^{+0.009910}_{-0.009910}$

7 Model-2

By following the same techniques as in Model-I, here with the use of Eqs. (28) and (39), We obtained the Hubble parameter $H(z)$ in terms of redshift z as,

$$H(z) = H_0(1+z)^{\beta} e^{\frac{\beta_0}{1+z}} \tag{47}$$

where $\beta = \frac{3(1+\omega_0+\omega_1)}{\alpha-1}$ and $\beta_0 = \frac{3\omega_1}{\alpha-1}$ are the free model parameters.

We have found the best fit estimated values of the parameters as $H_0 = 62.984292^{+0.415798}_{-0.415798}$, $\beta = 0.918487^{+0.009755}_{-0.009755}$ and $\beta_0 = 1.018775^{+0.009910}_{-0.009910}$ shown in Table 2 where as the theoretical curve of Hubble function (against the 57 measurements of $H(z)$ dataset) and distance modulus (against the Pantheon dataset) with ΛCDM model with their corresponding error bars is most likely seen in Figs. 10 and 11 and the likelihood contours for the same parameters at 1σ, 2σ level is shown in Fig. 12.

7.1 Cosmological parameters

7.1.1 The equation of state parameter

The graphical behaviour of the EoS parameter is depicted pretty clearly in Fig. 13. Our model's EoS parameter shows a shift in value with cosmic evolution, going from a positive to a negative, suggesting that the variable EoS parameter is what drives the transition from radiation-dominated matter to $-1 < \omega_\Lambda$ and $-1 > \omega_T$ which is the same as of our model-I. The model starts in a period where matter is predominate, varies in the quintessence area $-1 < \omega - \Lambda < \frac{1}{-3}$, and then moves closer to the phantom

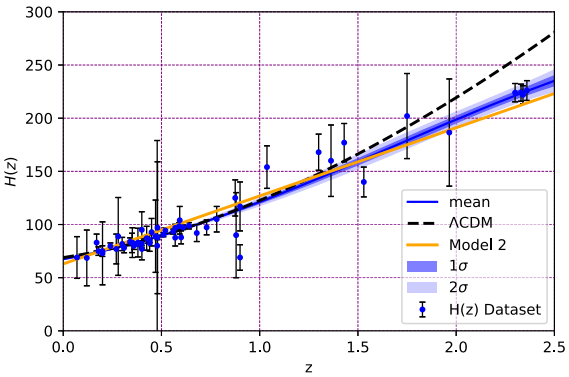


Fig. 10 The figure shows the theoretical curve of Hubble function $H(z)$ and ΛCDM models against the 57 measurements of H(z) datasets, shown in blue dots with their corresponding error bars

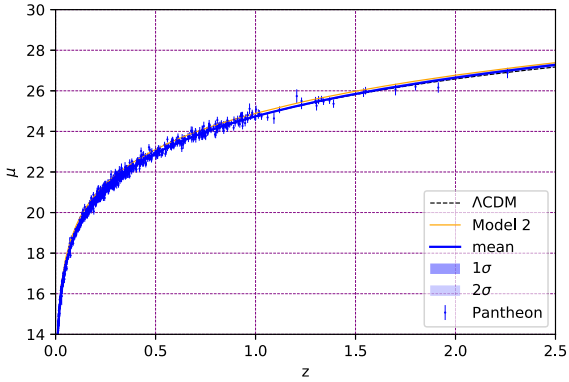


Fig. 11 The figure shows the theoretical curve of distance modulus $\mu(z)$ of our model and ΛCDM models against the Pantheon data, shown in pink dots with their corresponding error bars

region $\omega_\Lambda < -1$. The current value of the EoS parameter $\omega_\Lambda = -1$ indicates that the cosmos is currently dominated by the ΛCDM model.

Also, the reality that the present day values of the EoS parameter, $\omega_0 = -0.6152$ for the $H(z)+SC+unCorBAO$ dataset, are in precise accord with current Planck observational records is another surprising outcome of our version together with Aghanim et al. Provide the regulations at the equation of state parameter's value for DE in [75]. It is clear that our predictions of the equation of state parameter's contemporary value fall within the observational bounds indicated in the preceding reference. It is crucial to be aware that our findings for the destiny are identical to the observations made in [76] which is near to same as of model-I.

7.1.2 The deceleration parameter

Using Eq. (47) in Eq. (29), the deceleration parameter observed as

$$q = -1 - \frac{\beta_0}{z+1} + \beta \tag{48}$$

In Fig. 14, the redshift dependent deceleration parameter's behaviour is depicted. Initially as $z \longrightarrow \infty$ the deceleration parameter $q = 0.34563$. At preset where as

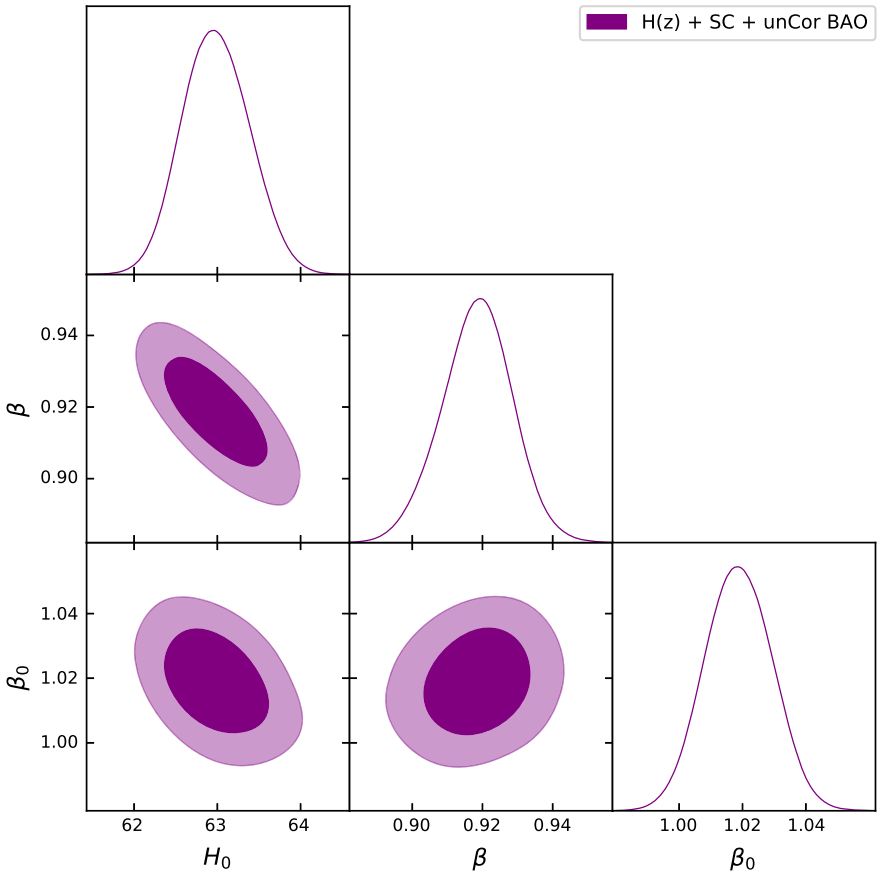


Fig. 12 This figure corresponds to 1σ and 2σ confidence contours obtained from $H(z)+SC+unCorBAO$ dataset obtained for the emergent model

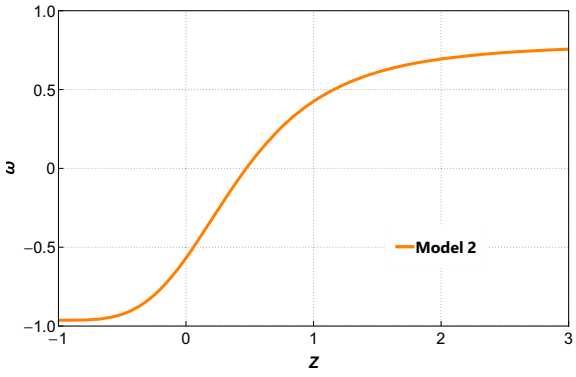


Fig. 13 Evolution of equation of state parameter with respect to redshift

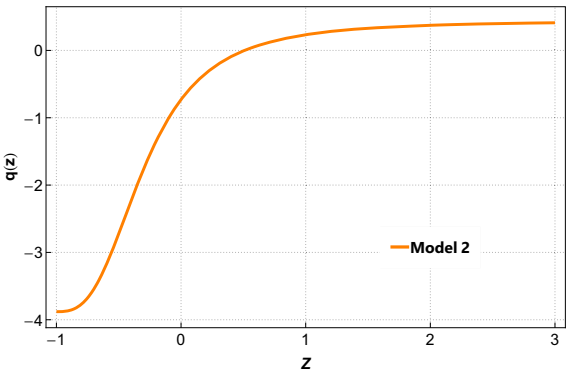


Fig. 14 Evolution of deceleration parameter with of Model 2 with respect to redshift

$z \longrightarrow 0$ the deceleration parameter is $q = -0.8254$ whereas in future at which $z \longrightarrow -1$ the value of the deceleration parameter is $q = -3.85343$. Hence the model seems to having a super-accelerated evolution.

The cosmos smoothly transitioned from the decelerated expansion to the accelerating expansion. Also, The contemporary behaviour flawlessly corresponds to the analysis of [74] probe the use of an anisotropic version and a decaying cosmological term for $n = 1.75$ equal as discused in model-I.

7.1.3 Statefinder diagnostics

Using Eq. (47) in Eq. (28), the statefinder diagnostics parameter observed as

$$r = 2\beta^2 - 3\beta - \frac{4(\beta-1)\beta_0}{z+1} + \frac{2\beta_0^2}{(z+1)^2} + 1 \tag{49}$$

$$s = \frac{4\beta_0^2 - 8(\beta-1)\beta_0(z+1) + 2\beta(2\beta-3)(z+1)^2}{3(z+1)((2\beta-3)(z+1) - 2\beta_0)} \tag{50}$$

On the contrary side, one could examine the profile of $\{r, s\}$ of Model 1 see Fig. 15, as Model 2 advances to the Chaplygin gas domain at late times by crossing intermediate ΛCDM fixed point $\{0, 1\}$ before quintessence domain. Meanwhile $\{r, q\}$ profile including further information on Model 2 as it evolves from the quintessence domain to Chaplygin gas (see Fig. 16)

7.1.4 $O_m(z)$ diagnostic

Using Eq. (39) in Eq. (29), the $O_m(z)$ diagnostic parameter observed as

$$O_m(z) = \frac{\frac{(z+1)^{2\beta} e^{\frac{2\beta_0}{z+1}}}{H_0^4} - 1}{(z+1)^3 - 1} \tag{51}$$

Figure 17 shows how the diagnostic parameter $O_m(z)$ behaves in relation to redshift z. The models are in the phantom domain because the $O_m(z)$ value is very small in

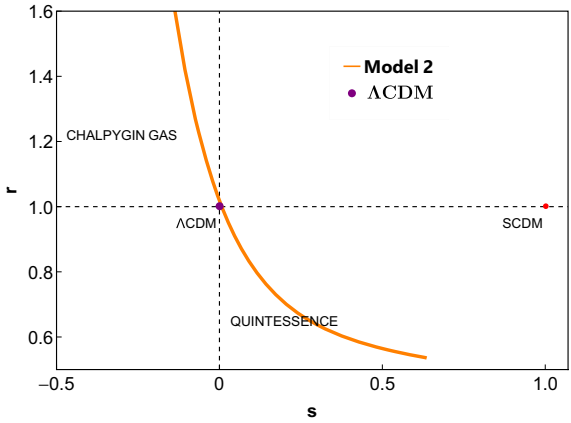


Fig. 15 Behavior of $\{r, s\}$ profile of Model 2 with respect to redshift

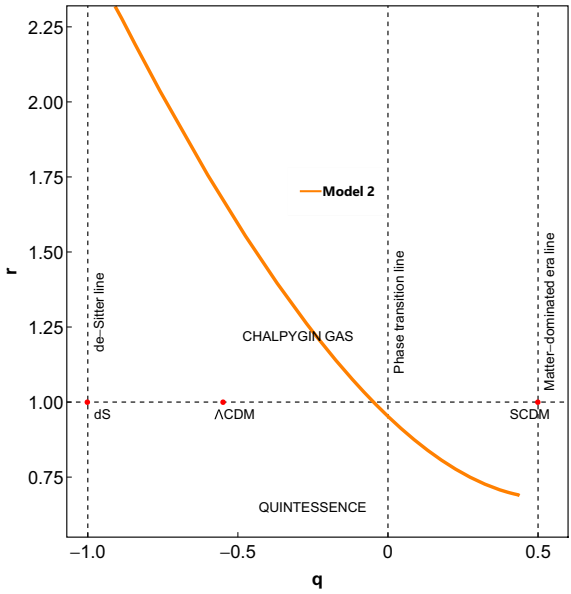


Fig. 16 Behavior of $\{r, q\}$ profile of Model 2 with respect to redshift

the redshift range of $z > 0$. The value increases and turns positive as z decreases, indicating that the model is the quintessence region (see Fig. 17).

7.1.5 Energy conditions

Using Eq. (39), the expressions of NEC, DEC and Strong energy condition respectively observed as,

$$NEC = \frac{1}{3}(-2)\alpha\beta k(z+1)^{\beta}\left(\alpha(z+1)^{\beta}+1\right)^{-2k-1}\left(H_0^2 m 6^n n(2n-3)\right.$$
$$\left.\left(H_0^2\left(\alpha(z+1)^{\beta}+1\right)^{-2k}\right)^{n-1}+6H_0^2\right), \tag{52}$$

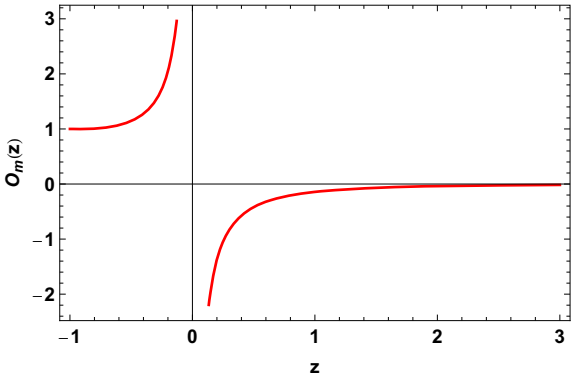


Fig. 17 The evolution of the $O_m(z)$ function of Model 2 with respect to redshift

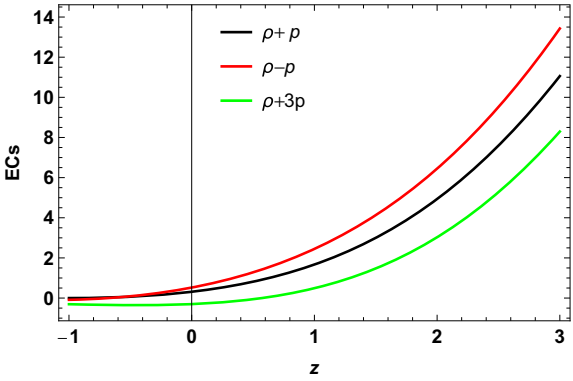


Fig. 18 Evolution of energy conditions with respect to redshift

$$DEC = \frac{2}{3}\left(\alpha(z+1)^{\beta}+1\right)^{-2k-1}\left(H_0^2 m 6^n\left(\alpha(z+1)^{\beta}(n(\beta k(2n-3)+6)-3)+6n-3\right)\right.$$
$$\left.\left(H_0^2\left(\alpha(z+1)^{\beta}+1\right)^{-2k}\right)^{n-1}+6H_0^2\left(\alpha(\beta k+3)(z+1)^{\beta}+3\right)\right), \quad (53)$$

$$SEC = -2\left(\alpha(z+1)^{\beta}+1\right)^{-2k-1}\left(H_0^2 m 6^n\left(\alpha(z+1)^{\beta}(n(\beta k(2n-3)+2)-1)+2n-1\right)\right.$$
$$\left.\left(H_0^2\left(\alpha(z+1)^{\beta}+1\right)^{-2k}\right)^{n-1}+6H_0^2\left(\alpha(\beta k+1)(z+1)^{\beta}+1\right)\right), \quad (54)$$

The behaviour of the Null (black curve), Dominant (red curve), and Strong (green curve) energy conditions of the present cosmological model is shown in the accompanying Fig. 18. The figure demonstrated that $\rho_\Lambda + p_\Lambda$, $\rho_\Lambda - p_\Lambda$, and $\rho_\Lambda + 3p_\Lambda$, indicating SEC is violated whereas NEC and DEC are energy conditions that are verified (Table 3).

8 Information criteria

To discuss sustainable model analysis, it is essential to understand the concept of information criteria (IC). Among the various ICs, the Akaike information criteria (AIC) [77] is widely used. The AIC is an approximation of the minimization of Kullback–

Table 3 Summary of the χ^2_{min}, χ^2_{red}, AIC and ΔAIC

Model	χ^2_{min}	χ^2_{red}	AIC	ΔAIC
ΛCDM Model	1094.32	0.972	1100.34	0
Model 1	1093.62	0.952	1101.66	1.314
Model 2	1096.06	0.954	1102.08	1.74

Leibler information and serves as an asymptotically unbiased estimator of Kullback–Leibler information. The AIC's Gaussian estimator can be expressed as [78–81]:

$$\mathrm{AIC} = -2\ln(\mathcal{L}_{\max}) + 2\kappa + \frac{2\kappa(\kappa+1)}{N-\kappa-1}, \tag{55}$$

where $\mathcal{L}_{\max}$ is the maximum likelihood function, κ is the total number of model parameters, and N is the total number of data points used for fitting the models to the data. In cases where $N \gg 1$, the formula simplifies to the original AIC expression as $\mathrm{AIC} = -2\ln(\mathcal{L}_{\max}) + 2\kappa$. The relative differences in IC values are compared when multiple models are considered. The difference in AIC values is given by $\triangle\mathrm{AIC} = \mathrm{AIC}_{\mathrm{Model}} - \mathrm{AIC}_{\min} = \triangle\chi^2_{\min} + 2\triangle\kappa$. In data analysis studies, a more favorable range for $\triangle$AIC is (0, 2). A less favorable range is (4, 7), while $\triangle\mathrm{AIC} > 10$ indicates weak support for the model.

9 Age of universe

To calculate the age of the universe based on the given models in Eqs. 39 and 47, we need to perform an integral with respect to redshift (z) from 0 to ∞. The age of the universe, denoted as t_0, is given by the integral:

$$t_0 = \int_0^\infty \frac{dz}{(1+z)H(z)} \tag{56}$$

Now, let's substitute the expression for $H(z)$ from Eqs. 39 and 47 into the integral (Table 4):

$$t_0 = \int_0^\infty \frac{dz}{(1+z)\left(\frac{H_0}{[1+\alpha(1+z)^\beta]^k}\right)} \tag{57}$$

and

$$t_0 = \int_0^\infty \frac{dz}{(1+z)\left(H_0(1+z)^\beta e^{\frac{\beta_0}{1+z}}\right)} \tag{58}$$

Table 4 Summary of the estimated age of the universe with uncertainty (standard deviation)

Model	Estimated age	Uncertainty
Model 1	13.49 billion years	1.51 billion years
Model 2	14.77 billion years	0.69 billions years

We also consider the uncertainties in the parameters to get an estimate range for the age of the universe. To perform the numerical integration with the given parameter values and uncertainties, we'll use the Monte Carlo method, which takes into account the uncertainties and provides a more realistic range for the age of the universe. These are the nessenary steps taken into account while for set up a Monte Carlo simulation to calculate the age of the universe:

- Sample random values for each of the parameters from their respective uncertainties following a normal distribution.
- For each set of sampled parameter values, calculate the value of $H(z)$ at various redshifts.
- Perform the integral over the range of redshifts using numerical methods like the trapezoidal rule or Simpson's rule.
- Repeat steps 1 to 3 for a large number of iterations to build a distribution of age values.
- Obtain the mean and standard deviation of the age distribution, which will provide an estimate of the age of the universe and its uncertainty.

10 Conclusions

In this study, we explored observational constraints of effective equation of state with $f(T)$ gravity using the flat FRW metric by showing the statistical viability of some $f(T)$ gravity models with the essential cosmological datasets. The values of the constraint parameters of model-I and model-II are reported in Tables 1 and 2. Our examination of the entire set of trustworthy $H(z)$ data at intermediate redshift $0.07 \leq z \leq 2.36$, along with the used datasets, however, provides some indications that the universe may be transitioning from a deceleration to an acceleration phase in the course of its evolution. We concur with the findings of other research, which assert that the simplest cosmological models, such ΛCDM, are consistent with our investigations evidence.

Also, first, we consider comparing Model 1 and Model 2 with the Hubble 57 data points, and with the ΛCDM model. The results of the comparison are presented in Figs. 1 and 10 and observe one from the figs that the model describes the Hubble data very well. Second, proceed to compare $\mu(z)$ distance modulus function of Model 1 and 2 with the Pantheon data shown in Figs. 2 and 11, one can observed that Model 1 and 2 fits the Pantheon distance modulus, based on 1048 observation points given by $0.01012 \leq z \leq 2.26$, very well. Some others key observations are as follows:

- The best-estimated values for MCMC results of the model-1 H_0, α, k, and β are summarized in Table 1 and model-2 parameters H_0, β and β_0 and summarized in Table 2.
- Figures 1, 2, 10 and 11 show the $H(z)$ and $\mu(z)$ and for model-1 amd model-2. Our derived models comparison with the ΛCDM model, along with error bars for the 57 points and 1048 of the considered Hubble and Pantheon datasets, respectively.
- The limits on the deceleration parameter $q(z)$ clearly show that the cosmic acceleration is predicted by the astronomical observations of $H(z)$ and SNe-Ia within the framework of $f(T)$ gravity cosmology. As a result, $f(T)$ cosmological models can adequately capture the universe's observable acceleration. Models 1 and 2 both show a phase change from deceleration to acceleration.
- For the consistency of the model we have used some observational datasets like $H(z) + SC + unCorBAO$ dataset. Figures 3 and 12 shows the plots of combined dataset samples of $H(z) + SC + unCorBAO$. We have obtained the likelihood contours for the model-1 and model-2 parameters.
- $O_m(z)$ diagnostic is also an effective method to discriminate various DE models from ΛCDM model according to the slope variation of $O_m(z)$. Figures 8 and 17 shows the behaviour of $O_m(z)$ versus z for model-1 and model-2. Our both models lies in quintessence region ($\omega > -1$).
- In order to compare the produced models with the conventional ΛCDM model and the SCDM model, the statefinder diagnostics are also supplied. The model behaviour is displayed in graphs in comparison with the standard ΛCDM model (see Figs. 6, 7, 15, 16).
- In Figs. 9 and 18 depicts the plots of energy conditions. Which reveal that NEC and DEC are satisfied, but SEC is not satisfied in both models. Due to the presence of exotic matter in the universe, there is a disagreement of SEC corresponds to the accelerating cosmic expansion of the universe.
- Figures 4 and 13 depicts the evolution of the effective equation of state parameter ω_Λ versus z. At present, our study reveals that ω_Λ approaches the ΛCDM value. Using the $H(z) + SC + unCorBAO$ dataset, the equation of state parameter $\omega(z)$ lies in the quintessence area, for both models which describes the characteristics of dark energy.

Data availability There are no new data associated with this article

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Riess, A.G., et al.: Astron. J. **116**, 1009 (1998)
2. Perlmutter, S., et al.: Astrophys. J. **517**, 565 (1999)
3. Nojiri, S.I., Odintsov, S.D.: Phys. Rep. **505**, 59 (2011)
4. Shamir, M.F.: Astrophys. Space Sci. **330**, 183 (2010)
5. Chirde, V.R., Shekh, S.H.: Bulg. J. Phys. **43**, 156 (2016)

6. Harko, T., Lobo, F.S.N., Nojiri, S., Odintsov, S.D.: Phys. Rev. D **84**, 024020 (2011)
7. Chirde, V.R., Shekh, S.H.: Astrophysics **58**, 106 (2015)
8. Chirde, V.R., Shekh, S.H.: J. Astrophys. Astron. **37**, 15 (2016)
9. Koussour, M., Bennai, M.: Int. J. Geom. Methods Mod. Phys. **19**, 2250038 (2022)
10. Koussour, M., Bennai, M.: Afrika Matematika **33**, 1 (2022)
11. Nojiri, S., Odintsov, S.D.: J. Phys. Conf. Ser. **66**, 012005 (2007)
12. Shekh, S.H., Arora, S., Chirde, V.R., Sahoo, P.K.: Int. J. Geom. Methods Mod. Phys. **17**, 2050048 (2020)
13. Lohakare, S.V., Tripathy, S.K., Mishra, B.: Physica Scripta **96**, 125039 (2021)
14. Shekh, S.H.: New Astron. **83**, 101464 (2021)
15. Shekh, S.H., Chirde, V.R., Sahoo, P.K.: Commun. Theor. Phys. **72**, 085402 (2020)
16. Wu, P., Yu, H.: Phys. Lett. B **693**, 415 (2010)
17. Hayashi, K., Shirafuji, T.: Phys. Rev. D **24**, 3312 (1981)
18. Maluf, J.W.: J. Math. Phys. **35**, 335 (1994)
19. Karami, K., Abdolmaleki, A.: Res. Astron. Astrophys. **137**, 757 (2013)
20. Amir, M.J., Yussouf, M.: Int. J. Theor. Phys. **54**, 2798 (2015)
21. Bamba, K., Geng, C.Q., Lee, C.C., Luo, L.W.: Astropart. Phys. **2011**(01), 021 (2011)
22. Chen, Y., Fay, S., Wang, Q.: J. Interact. Market. **25**, 85 (2011)
23. Amanullah, R., et al.: Astrophys. J. **716**, 712 (2010)
24. Nojiri, S., Odintsov, S.D., Tsujikawa, S.: Phys. Rev. D **71**, 063004 (2005)
25. Caldwell, R.R., Kamionkowski, M., Weinberg, N.N.: Phys. Rev. Lett. **91**, 071301 (2003)
26. Barrow, J.D.: Class. Quantum Gravity **21**, L79 (2004)
27. Abadji, H.F., Ganiou, M., Houndjo, M.J.S., Tossa J. (2019) arXiv preprint arXiv:1905.00718
28. Akarsu, O., Dereli, T., Kumar, S.: Eur. Phys. J. Plus **129**, 1 (2014)
29. Bamba, K.: Phys. Rev. D **89**, 083518 (2014)
30. Capozziello, S., Farooq, O., Luongo, O.: Phys. Rev. D **90**, 044016 (2014)
31. Guo, R.Y., Zhang, X.: Eur. Phys. J. C **76**, 163 (2016)
32. Verde, L., Protopapas, P., Jimenez, R.: Phys. Dark Univ. **5**, 307 (2014)
33. Chen, Y., Kumar, S., Ratra, B.: Astrophys. J. **835**, 89 (2017)
34. Wilson, W., Farooq, M.O.: J. Appl. Math. Comput. **1**, 1 (2017)
35. Farooq, O., Madiyar, F.R., Crandall, S., Ratra, B.: Astrophys. J. **835**, 26 (2017)
36. Goswami, G.K., Yadav, A.K., Mishra, B., Tripathy, S.K.: Fortschritte der Physik **69**, 2100007 (2021)
37. Shrivastava, P., Khan, A.J., Goswami, G.K, Yadav, A.K., Singh J.K.: (2021). arXiv preprint arXiv:2107.05044 (2021)
38. Scolnic, D.M., et al.: Astrophys. J. **859**, 101 (2018)
39. Marra, V., Amendola, L., Sawicki, I., Valkenburg, W.: Phys. Rev. Lett. **110**, 241305 (2013)
40. Mukherjee, A.: Mon. Not. R. Astron. Soc. **460**, 273 (2016)
41. Mamon, A.A., Das, S.: Eur. Phys. J. C **77**, 495 (2017)
42. Sahni, V., et al.: JETP Lett. **77**, 201 (2003)
43. Shahalam, M., Sami, S., Agarwal, A.: Mon. Not. R. Astron. Soc. **448**, 2948 (2015)
44. Sharma, U.K., Srivastava, S.: Mod. Phys. Lett. A **35**, 2050318 (2020)
45. Zunckel, C., Clarkson, C.: Phys. Rev. Lett. **101**, 181301 (2008)
46. Bouali, A., et al.: Fortschritte der Physik 2300033 (2023)
47. Bouali, A., et al.: arXiv preprint arXiv:2304.13137 (2023)
48. Bouali, A., et al.: arXiv preprint arXiv:2304.02652 (2023)
49. Bouali, A., et al.: arXiv preprint arXiv:2301.12107 (2023)
50. Bouali, A., Shukla, B.K., Chaudhary, H., Tiwari, R.K.: Int. J. Geom. Methods Mod. Phys. 2350152 (2023)
51. Arora, D., Pacif, S.K.J.: arXiv preprint arXiv:2304.09749 (2023)
52. Linton, M.S., et al.: J. Cosmol. Astropart. Phys. **04**, 043 (2018)
53. Capozziello, S., Nojiri, S.I., Odintsov, S.D.: Phys. Lett. B **781**, 99 (2018)
54. Alvarenga, F.G., Houndjo, M.J.S., Monwanou, A.V., Orou, J.B.C.: Int. J. Mod. Phys. **4**, 130 (2013)
55. Liu, D., Reboucas, M.J.: Phys. Rev. D **86**, 083515 (2012)
56. Bamba, K., Ilyas, M., Bhatti, M.Z., Yousaf, Z.: Gen. Relativ. Gravit. **49**, 1 (2017)
57. Atazadeh, K., Darabi, F.: Gen. Relativ. Gravit. **46**, 1 (2014)
58. Mishra, B., Esmeili, F.M., Ray, S.: Indian J. Phys. **95**, 2245 (2021)
59. Anderson, L., et al.: Mon. Not. R. Astron. Soc. **441**, 24 (2014)

60. Alam, U., Bag, S., Sahni, V.: Phys. Rev. D **95**, 023524 (2017)
61. Beutler, F., et al.: Mon. Not. R. Astron. Soc. **416**, 3017 (2011)
62. Scolnic, D.M., et al.: Astrophys. J. **859**, 101 (2018)
63. Bouali, A., et al.: Eur. Phys. J. C **83**(2), 121 (2023)
64. Gaztanaga, E., Cabre, A., Hui, L.: Mon. Not. R. Astron. Soc. **399**, 1663 (2009)
65. Kowalski, M., et al.: Supernova cosmology project. Astrophys. J. **686**, 749 (2008)
66. Amanullah, R., Lidman, C., Rubin, D., Aldering, G., Astier, P., Barbary, K., Burns, M.S., Conley, A., Dawson, K.S., Deustua, S.E., et al.: Astrophys. J. **716**, 712 (2010)
67. Suzuki, N., et al.: Supernova cosmology project. Astrophys. J. **746**, 85 (2012)
68. Betoule, M., et al.: SDSS Astron. Astrophys. **568**, A22 (2014)
69. Scolnic, D.M., et al.: Pan-STARRS1. Astrophys. J. **859**, 101 (2018)
70. Risaliti, G., Lusso, E.: Astrophys. J. **815**, 33 (2015). arXiv:1505.07118 [astro-ph.CO]
71. Blake, C., et al.: Mon. Not. R. Astron. Soc. **425**, 405 (2012). arXiv:1204.3674 [astro-ph.CO]
72. Alam, S., et al.: BOSS. Mon. Not. R. Astron. Soc. **470**, 2617 (2017). arXiv:1607.03155 [astro-ph.CO]
73. Benisty, D., Staicova, D.: Testing late-time cosmic acceleration with uncorrelated baryon acoustic oscillation dataset. Astron. Astrophys. **647**, A38 (2021)
74. Tiwari, R.K., Beesham, A.: Astrophys. Space Sci. **363**, 234 (2018)
75. Aghanim, N., et al.: Astron. Astrophys. **641**, A6 (2020)
76. Shekh, S.H.: Phys. Dark Universe **33**, 100850 (2021)
77. Akaike, H.: IEEE Trans. Autom. Control **19**(6), 716 (1974)
78. Li, M., Li, X., Zhang, X.: Sci. China Phys. Mechanics. Astron. **53**(9), 1631 (2010)
79. Burnhan, K., Anderson, D.R.: Model Selection and Multimodel Inference. Springer, New York (2002)
80. Burnham, K.P., Anderson, D.R.: Sociol. Methods Res. **33**(2), 261–304 (2004)
81. Liddle, A.R.: Mon. Not. R. Astron. Soc. Lett. **377**(1), L74 (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Offensive Hinglish Text Classification Using Longformer, Bigbird & HingMBert

Yakshit Bansal
Department of Information Technology
Delhi Technological University
Delhi, India
yakshit.bansal2304@gmail.com

Sumit Samdarshi
Department of Information Technology
Delhi Technological University
Delhi, India
sumitsamdarshi_2k19it127@dtu.ac.in

Sushant Kumar
Department of Information Technology
Delhi Technological University
Delhi, India
sushantkumar_2k19it130@dtu.ac.in

Dinesh Kumar Vishwakarma*
Department of Information Technology
Delhi Technological University
Delhi, India
dinesh@dtu.ac.in

2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT) | 979-8-3503-3936-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/APSIT58554.2023.10201686

***Abstract*— With a sharp rise in the number of users and skilled fluency in the use of "Hinglish", a cross-lingual text of Hindi and English, in a linguistically diverse country, India, it has increasingly become important to analyze social content written in the aforementioned language over platforms such as Twitter, Reddit, Facebook. This project focuses on the classification of Hinglish text into offensive and non-offensive categories. We present a state-of-the-art novel use of "Longformer" and "Bigbird" transformer-based language modeling models on Hinglish texts for natural language processing (NLP) sentiment classification task, alongside mBERT, pre-trained for Hinglish, hence "HingMBert" model. Our dataset is a combination of pre-existing datasets and further addition of approximately 8,200 records, a total of 24,660 records. The tweepy library has been used to scrape novel tweets. Experimental results for Longformer, Bigbird, and HingMBert have achieved F1 scores of 0.76, 0.77, and 0.79, and accuracy of 76%, 77%, and 80% respectively. F1 score of 0.78 and an accuracy of 78% was achieved using the majority voting ensemble technique.**

Keywords—longformer, bigbird, hingmbert, hinglish

I. Introduction

As per the national statistic of India, approximately 10.6% of the Indian population can speak English, and approximately 45% of the population speaks Hindi. Hindi and English, both being the official languages of India, the current trend is population speaking English is growing more rapidly than the population speaking Hindi. Therefore, a blend of composite dialects is observed in either scenario called "Hinglish." In slang "Hinglish" is referred to as "WhatsApp Language"; meaning Hindi written using English script. Currently 350+ million users of social media in India use Hinglish for communication over social media platforms.

With the advent of a composite script in use, advantageous in accumulating vocabulary from multiple scripts like Punjabi, Marathi, Gujarati, etc., and with the sharp rise in the number of social media platforms such as Twitter, YouTube, Instagram, Facebook, TikTok, Reddit, Snapchat, the narrative of "Freedom of Speech" and ease of expressing views has risen exponentially. The recoil of this expression is defaming, trolling, offending the minorities, and expressing anger, the domino effect leads to attacking personalities or groups based on gender, caste, creed, race, ethnicity, ideologies, and LGBTQIA+ groups.

Henceforth, the use of offensive language, which includes hate speech and the use of abusive vocabulary is increasing exponentially. The task of analyzing such language and identifying a tone of expression is currently carried out throughout the NLP domain. Consequently, the task is divided into sectors where each problem statement tries to identify either one specific part or targets a domain related to a group, region, or circumstances. For example, [1] studies the consequences of hate speech leading to aggression, [2], studied the content that went on social media during the pandemic related to covid.

II. Related Work

Using a transformer-based approach for offensive/hate speech detection specific to code-mixed bilingual speakers of Hinglish has been limited. [3] used multilingual bidirectional encoder representations from transformers (mBERT) embedding on 4575 derived tweets over machine learning models, logistic regression (LR), support vector machine (SVM), k-nearest neighbors (KNN), random forest (RF), naïve bayes (NB). [1] used pseudo-labeling techniques along with BERT and long short term memory (LSTM). [4] also used pre-trained BERT, and one-hot encoding for an offensiveness detection system detection system. [5] experimented with RoBERTa-base, DistilBERT-Base, and BERT-base models along with LR, NB-SVM ensemble, for Hinglish sentiment detection.

For NLP-driven summarization, long document classification, [6] introduces the linearly increasing attention span mechanism, [7] introduces the sparse attention mechanism. So far, Bigbird and Longformer have been majorly used in clinical records classification. [8] successfully carried out NLP downstream tasks on clinical records, where the named Clinical-Longformer and Clinical-Bigbird out-performed BERT. [9] used hierarchical transformer network including Bigbird. [10] used Bigbird, while [11] used Longformer for disease classification while maintaining the integrity of the Specifications

III. Dataset

The dataset has been fine-tuned on 4 datasets. 4579 records from [12] 12,000 records from [1], and 8,200 records were scraped from Twitter using the "Tweepy library" and the using the set of offensive keywords derived from "Hatebase.org". For non-offensive tweets, a similar strategy was used.

All the derived datasets have been manually cross-verified for use and labeling was done on the records fetched from Twitter. The scraped record contained both classes of

979-8-3503-3936-9/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:21 UTC from IEEE Xplore. Restrictions apply.

tweets and the skewness of the data was appropriately removed.

Pretraining data: Transformer-based language models need a huge corpus for pretraining. L3Cube-HingCorpus [13] has been used for the pretraining of Longformer, Bigbird, and HingMBert. The corpus contains approximately 100,000 sentences for training.

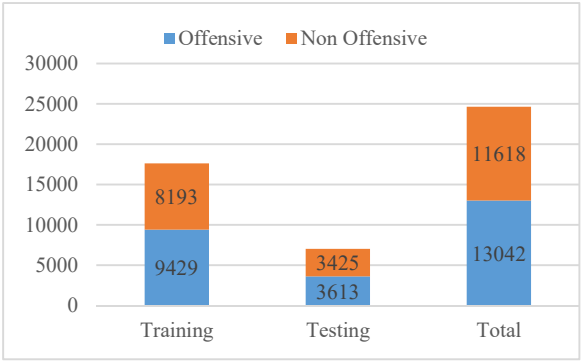


Fig.1. Comparative data for Offensive and Non-offensive records.

IV. Preprocessing

The data derived from the fore-mentioned resources contained multiple irregularities, for example, '#hashtags', 'URLs', and '@tagging', plus the tweets fetched had emojis, punctuations, repetitive characters, etc.

All the words were converted to lowercase, tokenization of the sentences was done, and to a feasible extent, lemmatization for English as well as for Hindi words was done.

The dataset was preprocessed for English as well as for Hindi stop words. Emojis were removed based on the ASCII code. Punctuations, URLs, and hashtags were removed too. Most commonly used spellings of Hinglish offensive words were already given in https://hatebase.org/, other inconsistencies were examined and alterations were less than 1 in 3000, which were altered to the most common version of the word.

V. Methodology

Self-attention mechanism differs in all three transformer models used. The difference lies in the attention span, and the said difference optimizes the space and time complexities of the algorithms. HingMBert has a full attention mechanism making it O(n^2), where n is the sequence length, space complexity transformer model. The model relates and creates a matrix for the entire document. Longformer uses global and sliding window attention mechanisms. The sliding window consists of 512 tokens and the global window spans across some of the special tokens of the sequence. Since the number of special tokens is very less than the entire sequence length of n, the access time complexity of the Longformer drops to linear, O(n). Bigbird includes random attention in addition to the window and global attention. To increase the access to an even larger sequence up to the length of 4096, i.e. making the quadratic access memory time of Longformer linear. The reason is the introduction of random attention. Random attention has logarithmic access to the tokens, i.e. instead of each token referring to every token of the sequence and utilizing quadratic memory, the random connection within the same layer of the network accesses only a few nodes and then directs to the destination node. On average each node or token is accessed and the average memory complexity reduces to linear as well.

Therefore, in our experimentation, HingMBert utilized the maximum memory of 700 MB, followed by Longformer 500 MB and Bigbird 400MB range. The time consumed in running these algorithms followed the same sequence.

As shown in Fig. 2, the different types of attention span used by the transformer models and at the end the major voting ensemble aggregating the result.

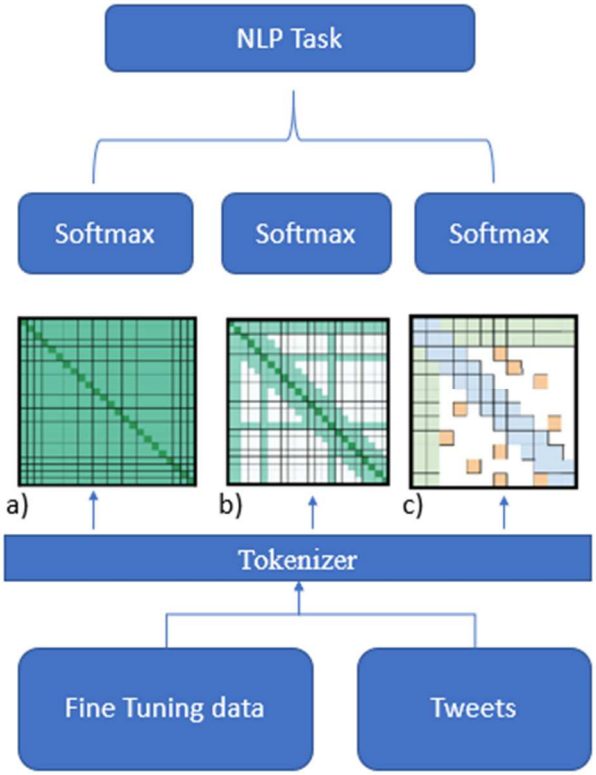


Fig.2. Working methodology. a) HingMBert full attention span [6] b) Longformer Global + sliding window attention [6] and c) Bigbird: Random + sliding + global attention span mechanisms [7].

VI. Experimental Setup

For Longformer and Bigbird model experimentation, the models were accessed from www.huggingface.co, and the HingMBert model was reused from L3Cube [13].

The pretraining of the models, unsupervised learning, was done on L3Cube-HingCorpus [13]. The models were imported using the "simpletransformers" library.

In the next step, the dataset was randomly shuffled and was divided into 75% and 25% ratios, for training and test purposes respectively. Thus, 17,622 records for training and 7,038 records for testing respectively were stored in different datasets.

Fine-tuning of all the models, Longformer, Bigbird, and HingMBert was done over 75% dataset in a batch size of 16 and the number of epochs was set to 8. The learning rate for all the models was set to the default value of 4e-5.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:21 UTC from IEEE Xplore. Restrictions apply.

VII. Results

The results obtained from the experimentation have been presented below. Accuracy, Precision, Recall, and F1-Score have been calculated for all the models used.

The classification matrix of the fine-tuned models of Longformer, Bigbird, and HingMBert are presented in Tables 2, 3, and 4 respectively. The classification matrix of majority voting is shown in Table 5.

TABLE I. Classification matrix for Longformer

	Predicted			
Actual	Label	0	1	All
	0	2602	889	3491
	1	823	2724	3547
	All	3425	3613	7038

TABLE II. Classification matrix for Bigbird

	Predicted			
Actual	Label	0	1	All
	0	2579	912	3491
	1	739	2808	3547
	All	3318	3720	7038

TABLE III. Classification matrix for hingMbert

	Predicted			
Actual	Label	0	1	All
	0	2851	640	3491
	1	767	2780	3547
	All	3618	3420	7038

TABLE IV. Classification matrix for major voting ensemble

	Predicted			
Actual	Label	0	1	All
	0	2740	751	3491
	1	755	2792	3547
	All	3495	3543	7038

TABLE V. Classification report of language models and majority voting ensemble

Model	Accuracy	Precision	Recall	F1 Score
Longformer	0.756749	0.753944	0.767973	0.760894
Bigbird	0.765416	0.754839	0.791655	0.772809
HingMBert	0.800085	0.812865	0.783761	0.798048
Majority Voting Ensemble	0.786019	0.788033	0.787144	0.787588

HingMBert has shown to have a maximum accuracy of 80%, followed by Bigbird 76.5% and Longformer at 75.7%. The above order has also been observed for F1-Score with 0.79, 0.77, and 0.76 by HingMBert, Bigbird, and Longformer respectively. Majority voting has produced resulted in an accuracy of 78.6% and F1-Score of 0.78.

The reason for increased accuracy in HingMBert is the full attention mechanism, which provided a better estimation of the sequences related to the pre-trained data than Longformer and Bigbird. Bigbird slightly outperformed Longformer because of the number of longer sequences which per predicted more accurately by Bigbird.

VIII. Conclusion and Future Work

This paper presents a state-of-the-art novel use of Longformer and Bigbird, a transformer-based language modeling model, alongside HingMBert. We concluded that currently, HingMBert has performed better than the rest. This study opens the door for the application of other transformer-based language modeling models, that have currently been used for English NLP sentiment analysis and other downstream processes. Possible prospects of this application include the use of multinomial sentiment analysis, which will also be enhanced by the use of larger datasets and increased pretraining corpus.

Acknowledgment

The authors would like to thank the Department of Information Technology, Delhi Technological University, Delhi, for providing constant support in terms of resources, necessary guidance, and constant motivation.

References

[1] A. Sengupta, S. K. Bhattacharjee, M. S. Akhtar, and T. Chakraborty, "Does aggression lead to hate? Detecting and reasoning offensive traits in hinglish code-mixed texts," *Neurocomputing*, vol. 488, pp. 598–617, Jun. 2022, doi: 10.1016/J.NEUCOM.2021.11.053.

[2] N. Vishwamitra, R. R. Hu, F. Luo, L. Cheng, M. Costello, and Y. Yang, "On Analyzing COVID-19-related Hate Speech Using BERT Attention," *Proceedings - 19th IEEE International Conference on Machine Learning and Applications, ICMLA 2020*, pp. 669–676, Dec. 2020, doi: 10.1109/ICMLA51294.2020.00111.

[3] S. Biradar, S. Saumya, and A. chauhan, "Fighting hate speech from bilingual hinglish speaker's perspective, a transformer- and translation-based approach.," *Soc Netw Anal Min*, vol. 12, no. 1, pp. 1–10, Dec. 2022, doi: 10.1007/S13278-022-00920-W/TABLES/7.

[4] S. H. Vissamsetti, K. S. S. Pravallika, S. C. Jadhav, and H. B. Kommanti, "Offensiveness Detection in Hinglish Code-Switched Language," *Communications in Computer and Information Science*, vol. 1631, pp. 124–135, 2022, doi: 10.1007/978-3-031-15556-7_9/COVER.

[5] M. Bhange, V. Io Bengaluru, and N. Kasliwal, "HinglishNLP at SemEval-2020 Task 9: Fine-tuned Language Models for Hinglish Sentiment Detection," *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pp. 934–939, 2020, doi: 10.18653/V1/2020.SEMEVAL-1.119.

[6] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The Long-Document Transformer," Apr. 2020, Accessed: May 08, 2023. [Online]. Available: https://arxiv.org/abs/2004.05150v2

[7] M. Zaheer *et al.*, "Big Bird: Transformers for Longer Sequences," *Adv Neural Inf Process Syst*, vol. 2020-December, Jul. 2020, Accessed: May 08, 2023. [Online]. Available: https://arxiv.org/abs/2007.14062v2

[8] Y. Li, R. M. Wehbe, F. S. Ahmad, H. Wang, and Y. Luo, "Clinical-Longformer and Clinical-BigBird: Transformers for long clinical sequences," Jan. 2022, Accessed: May 08, 2023. [Online]. Available: https://arxiv.org/abs/2201.11838v3

[9] Y. Si and K. Roberts, "Three-level Hierarchical Transformer Networks for Long-sequence and Multiple Clinical Documents Classification," Apr. 2021, Accessed: May 08, 2023. [Online]. Available: https://arxiv.org/abs/2104.08444v2

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:21 UTC from IEEE Xplore. Restrictions apply.

[10] G. Michalopoulos, M. Malyska, N. Sahar, A. Wong, and H. Chen, "ICDBigBird: A Contextual Embedding Model for ICD Code Classification," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 330–336, Apr. 2022, doi: 10.18653/v1/2022.bionlp-1.32.

[11] Q. Chen and Y. Hong, "Longformer: Longitudinal Transformer for Alzheimer's Disease Classification with Structural MRIs," Feb. 2023, Accessed: May 08, 2023. [Online]. Available: https://arxiv.org/abs/2302.00901v2

[12] P. Mathur, R. Sawhney, M. Ayyar, and R. R. Shah, "Did you offend me? Classification of Offensive Tweets in Hinglish Language," *2nd Workshop on Abusive Language Online - Proceedings of the Workshop, co-located with EMNLP 2018*, pp. 138–148, 2018, doi: 10.18653/V1/W18-5118.

[13] R. Joshi, "L3Cube-HindBERT and DevBERT: Pre-Trained BERT Transformer models for Devanagari based Hindi and Marathi Languages," Nov. 2022, Accessed: May 08, 2023. [Online]. Available: https://arxiv.org/abs/2211.11418v4

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:21 UTC from IEEE Xplore. Restrictions apply.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/362429550

Starlike Functions associated with a Non-Ma-Minda Function

Preprint · August 2022

CITATIONS
0

READS
66

2 authors:

s. Sivaprasad Kumar
Delhi Technological University
97 PUBLICATIONS **909** CITATIONS
SEE PROFILE

Pooja Yadav
Delhi Technological University
3 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Coefficient Estimates and Differential Subordinations for certain class of analytic functions View project

All content following this page was uploaded by Pooja Yadav on 30 December 2022.

The user has requested enhancement of the downloaded file.

arXiv:2208.01245v2 [math.CV] 9 Sep 2022

ON A CLASS OF CERTAIN NON-UNIVALENT FUNCTIONS

S. SIVAPRASAD KUMAR AND POOJA YADAV

Abstract. In this paper, we introduce a family of analytic functions given by

$$\psi_{A,B}(z) := \frac{1}{A-B}\log\frac{1+Az}{1+Bz},$$

which maps univalently the unit disk onto either elliptical or strip domains, where either $A=-B=\alpha$ or $A=\alpha e^{i\gamma}$ and $B=\alpha e^{-i\gamma}$ ($\alpha\in(0,1]$ and $\gamma\in(0,\pi/2]$). We study a class of non-univalent analytic functions defined by

$$\mathcal{F}[A,B] := \left\{ f\in\mathcal{A} : \left(\frac{zf'(z)}{f(z)}-1\right) \prec \psi_{A,B}(z)\right\}.$$

Further, we investigate various characteristic properties of $\psi_{A,B}(z)$ as well as functions in the class $\mathcal{F}[A,B]$ and obtain the sharp radius of starlikeness of order δ and univalence for the functions in $\mathcal{F}[A,B]$. Also, we find the sharp radii for functions in $\mathcal{BS}(\alpha) := \{f\in\mathcal{A} : zf'(z)/f(z)-1 \prec z/(1-\alpha z^2),\ \alpha\in(0,1)\}$, $\mathcal{S}_{cs}(\alpha) := \{f\in\mathcal{A} : zf'(z)/f(z)-1 \prec z/((1-z)(1+\alpha z)),\ \alpha\in(0,1)\}$ and others to be in the class $\mathcal{F}[A,B]$.

1. Introduction

Let $\mathcal{A}$ denote the class of analytic functions f defined on the open unit disk $\mathbb{D}=\{z\in\mathbb{C} : |z|<1\}$, such that $f(z)=z+a_2z^2+a_3z^3+\cdots$. The subclass of $\mathcal{A}$ consisting of all univalent functions is denoted by $\mathcal{S}$. For two analytic functions f and g, we say that f is subordinate($\prec$) to g, if $f(z)=g(\omega(z))$ where $\omega(z)$ is a Schwarz function. Let $\mathcal{S}^*$ be the class of starlike functions, consists of functions $f\in\mathcal{S}$ such that $zf'(z)/f(z)\prec(1+z)/(1-z)$. In [6], Ma and Minda unifies all the subclasses of $\mathcal{S}^*$ as follows:

$$\mathcal{S}^*(\Phi) := \left\{ f\in\mathcal{S} : \frac{zf'(z)}{f(z)} \prec \Phi(z)\right\}, \tag{1.1}$$

where Φ is univalent with $\Phi'(0)>0$ satisfying

A. $\Phi(z)\prec(1+z)/(1-z)$
B. $\Phi(\mathbb{D})$ is symmetric about real axis and starlike with respect to $\Phi(0)=1$.

Further, Kumar and Banga [3], called such Φ as *Ma-Minda function* and the class of all Ma-Minda functions is denoted by $\mathcal{M}$. In [3], authors extensively studied Ma-Minda functions and classified them as *non-Ma-Minda function of type-A* whenever condition A doesn't hold, the class of all such functions is denoted by $\widetilde{\mathcal{M}}_A$. In past many authors studied the class (1.1) for various choices of $\Phi\in\mathcal{M}$, such as $\mathcal{S}^*(2/(1+e^{-z})) =: \mathcal{S}^*_{SG}$ [1], $\mathcal{S}^*(e^z) =: \mathcal{S}^*_e$ [9], $\mathcal{S}^*(z+\sqrt{1+z^2})$ [10], $\mathcal{S}^*(\sqrt{1+z}) =: \mathcal{S}^*_L$ [13], etc. Also the well known classes $\mathcal{S}^*(A,B)$ and $\mathcal{SS}^*(\alpha)$ are also obtained by taking $\Phi(z)(\in\mathcal{M})$ as $(1+Az)/(1+Bz)$ $(-1\le B<A\le 1)$ and $((1+z)/(1-z))^\alpha$ $(\alpha\in(0,1])$ respectively. In particular, $\mathcal{S}^*(1,-1) =: \mathcal{S}^*$, $\mathcal{S}^*(1-2\delta,-1) =: \mathcal{S}^*(\delta)$ $(\delta\in[0,1))$, the class of starlike functions of order δ. Robertson [11] introduced and investigated the class of starlike

2010 *Mathematics Subject Classification.* 30C45, 30C50, 30C80.

Key words and phrases. Analytic, Subordination, Logatithmic function, Starlike function, Non-univalent function.

The second author is supported by The Council of Scientific and Industrial Research(CSIR). Ref.No.:08/133(0030)/2019-EMR-I.

functions of order δ, denoted by $\mathcal{S}^*_\delta$ for $\delta \leq 1$. Note that if $\delta < 0$, then the functions in $\mathcal{S}^*_\delta$ may not be univalent, i.e. if $\delta < 0$, then $\Phi(z) = (1 + (1 - 2\delta)z)/(1 - z) \in \widetilde{\mathcal{M}}_A$ and $\mathcal{S}^*_\delta \not\subset \mathcal{S}^*$. Later many authors considered the classes associated with non-Ma-Minda functions of type-A such as the class by Uralegaddi [14]

$$\mathscr{M}(\beta) = \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} \prec \frac{1 + (2\beta - 1)z}{1 - z},\ \beta > 1 \right\},$$

the classes associated with the Booth Lemniscate and Cissoid of Diocles

$$\mathcal{BS}^*(\alpha) := \left\{ f \in \mathcal{A} : \left(\frac{zf'(z)}{f(z)} - 1 \right) \prec \frac{z}{1 - \alpha z^2},\ \alpha \in [0, 1) \right\},$$

$$\mathcal{S}^*_{cs}(\alpha) := \left\{ f \in \mathcal{A} : \left(\frac{zf'(z)}{f(z)} - 1 \right) \prec \frac{z}{(1 - z)(1 + \alpha z)},\ \alpha \in [0, 1) \right\},$$

which were introduced by [2] and [7] respectively. Similarly, Masih and Kanas [8] studied the class $\mathcal{ST}^*_L(s) := \{f \in \mathcal{A} : zf'(z)/f(z) \prec \mathbb{L}_s(z)\}$ associated with the Limaçon of Pascal $\mathbb{L}_s(z) = (1+sz)^2$, $s \in [-1, 1]\backslash\{0\}$. Note that $\mathbb{L}_s(z) \in \widetilde{\mathcal{M}}_A$, whenever $|s| > 1/\sqrt{2}$. Recently, Kumar and Gangania [5] studied the following class:

$$\mathcal{F}(\psi) := \left\{ f \in \mathcal{A} : \frac{zf'(z)}{f(z)} - 1 \prec \psi(z) \right\}, \tag{1.2}$$

where $\psi(z) \in \mathcal{S}^*$. If $1 + \psi(z) \prec (1 + z)/(1 - z)$ and $1 + \psi(\mathbb{D})$ is symmetric about the real axis with $\psi'(0) > 0$ then $\mathcal{F}(\psi)$ reduces to the class $\mathcal{S}^*(1+\psi)$, a Ma-Minda class and if $1+\psi(z) \not\prec (1+z)/(1-z)$ then the functions in $\mathcal{F}(\psi)$ may not be univalent, thus $\mathcal{F}(\psi) \not\subseteq \mathcal{S}^*$. Note that all above mentioned classes associated with non-Ma-Minda functions of type-A are the special cases of $\mathcal{F}(\psi)$ such as $\mathcal{F}((1 + (2\beta - 1)z)/(1 - z)) =: \mathscr{M}(\beta)$, $\mathcal{F}(z/(1 - \alpha z^2)) =: \mathcal{BS}^*(\alpha)$, $\mathcal{F}(z/((1 - z)(1 + \alpha z))) =: \mathcal{S}^*_{cs}(\alpha)$ and $\mathcal{F}((1 + sz)^2) =: \mathbb{L}_s(z)$.

Motivated by the above, we define a family of functions $\psi_{A,B}(z)$ as

$$\psi_{A,B}(z) := \frac{1}{A - B} \log \left(\frac{1 + Az}{1 + Bz} \right) \qquad z \in \mathbb{D}, \tag{1.3}$$

where either $A = -B = \alpha$ or $A = \alpha e^{i\gamma}$ and $B = \alpha e^{-i\gamma}$, for $0 < \alpha \leq 1$ and $0 < \gamma \leq \pi/2$. The function defined by (1.3) is analytic everywhere except the branchcuts $\{-\infty < \operatorname{Re} Az \leq -1,\ \operatorname{Im} Az = 0\} \cup \{-\infty < \operatorname{Re} Bz \leq -1,\ \operatorname{Im} Bz = 0\}$ with $\psi_{A,B}(0) = \psi'_{A,B}(0) - 1 = 0$, i.e. $\psi_{A,B}(z) \in \mathcal{A}$. Now we introduce the following classes associated with $\psi_{A,B}(z)$:

Definition 1.1. Let $p \in \mathcal{S}$. Then $p \in \mathcal{L}[A, B]$ if and only if

$$p(z) \prec \psi_{A,B}(z).$$

Definition 1.2. Let $f \in \mathcal{A}$. Then $f \in \mathcal{F}[A, B]$ if and only if

$$\frac{zf'(z)}{f(z)} - 1 \prec \psi_{A,B}(z). \tag{1.4}$$

In section 2 we investigate various characteristic properties of functions in the classes $\mathcal{L}[A, B]$ and $\mathcal{F}[A, B]$. We also obtain the extremal function of the class $\mathcal{F}[A, B]$ which is non-univalent and study its various geometric properties. Further, in section 3 we derive the sharp radius of starlikeness of order δ, univalence for the functions in $\mathcal{F}[A, B]$ and also find the sharp radii for functions in $\mathcal{BS}(\alpha)$, $\mathcal{S}_{cs}(\alpha)$ and others to be in the class $\mathcal{F}[A, B]$.

2. Characteristics of $\mathcal{L}[A,B]$ and $\mathcal{F}[A,B]$

In this section, we aim to find various geometric properties of $\psi_{A,B}(z)$ and obtain certain bounds and inclusion relations for functions in the class $\mathcal{L}[A,B]$. Further, we deduce the extremal function and derive the growth and covering theorems for the class $\mathcal{F}[A,B]$. Note that

$$\psi_{A,B}(z)=\sum_{n=1}^{\infty}(-1)^{n+1}C_n z^n,$$

where

$$C_n=\frac{A^n-B^n}{n(A-B)}=\frac{1}{n}\sum_{k=0}^{n-1}A^kB^{n-1-k}\qquad(n=1,2,\ldots). \tag{2.1}$$

Remark 2.1. From (2.1) we have

(i) Since C_n is real valued for all n, $\psi_{A,B}(\mathbb{D})$ is symmetric with respect to real axis.

(ii) $|C_n|=\left|\frac{1}{n}\sum_{k=0}^{n-1}A^kB^{n-1-k}\right|\leq\alpha^{n-1}\leq 1.$

Lemma 2.2. *The function $\psi_{A,B}(z)$ is convex and univalent on $\mathbb{D}$.*

Proof. Let $H(z):=1+z\psi''_{A,B}(z)/\psi'_{A,B}(z)$. By [12, Corollary 3], it is enough to show that $\operatorname{Re}H(z)>0$. Since $|A|=|B|=\alpha\leq 1$, using basic calculation with [4, Theorem 1], we obtain

$$\begin{aligned}\operatorname{Re}H(z)&=\operatorname{Re}\left(-1+\frac{1}{1+Az}+\frac{1}{1+Bz}\right)\\&>-1+\frac{2}{1-\alpha}>0.\end{aligned}$$

Thus the result holds. ∎

Theorem 2.3. *Let $p\in\mathcal{L}[A,B]$, then in the disk $\mathbb{D}_r=\{z\in\mathbb{C}:|z|\leq r<1\}$,*

(i) For $A=-B=\alpha$, we have

$$|\operatorname{Re}p(z)|\leq\frac{1}{2\alpha}\log\left(\frac{1+\alpha r}{1-\alpha r}\right) \tag{2.2}$$

$$|\operatorname{Im}p(z)|\leq\frac{1}{2\alpha}\sin^{-1}\left(\frac{2\alpha r}{1+\alpha^2r^2}\right). \tag{2.3}$$

(ii) For $A=\alpha e^{i\gamma}$ and $B=\alpha e^{-i\gamma}$, we have

$$\frac{-1}{2\alpha\sin\gamma}(\eta-\tau)\leq\operatorname{Re}p(z)\leq\frac{1}{2\alpha\sin\gamma}(\eta+\tau). \tag{2.4}$$

$$\frac{1}{2\alpha\sin\gamma}\log T_2\leq\operatorname{Im}p(z)\leq\frac{1}{2\alpha\sin\gamma}\log T_1. \tag{2.5}$$

where

$$\eta:=\sin^{-1}\left(\frac{2\alpha r\sin\gamma}{\sqrt{1+\alpha^4r^4-2\alpha^2r^2\cos 2\gamma}}\right),$$

$$\tau:=\tan^{-1}\left(\frac{-\alpha^2r^2\sin 2\gamma}{1-\alpha^2r^2\cos 2\gamma}\right)$$

and

$$T_j:=\left(\frac{\sqrt{1+\alpha^4r^4-2\alpha^2r^2\cos 2\gamma}+(-1)^j2\alpha r\sin\gamma}{1-\alpha^2r^2}\right)^{-1}\quad(j=1,2).$$

Proof. Let w be the Schwarz function such that $w(0) = 0$ and $|w(z)| \leq |z| = r < 1$ for all $z \in \mathbb{D}$. Define the function $F_w : \mathbb{D} \to \mathbb{C}$ such that $p(z) = \psi_{A,B}(w(z)) = \dfrac{1}{A-B}\log\left(F_w(z)\right)$. We have

$$F_w(z) := \frac{1+Aw(z)}{1+Bw(z)},$$

which maps unit disk into the disk

$$F_w(\mathbb{D}) \subset \left\{\zeta \in \mathbb{C}; \left|\zeta - \frac{1-A\overline{B}}{1-|B|^2}\right| < \frac{|A-B|}{1-|B|^2}\right\}.$$

As $|A| = \alpha \in (0,1]$, therefore origin $O \notin F_w(\mathbb{D})$ (for more details see [4]). Now using [4, Theorem 1] with $|z| = r < 1$, we have

$$\frac{|1-A\overline{B}r^2| - |A-B|r}{1-|B|^2r^2} \leq |F_w(z)| \leq \frac{|1-A\overline{B}r^2| + |A-B|r}{1-|B|^2r^2},$$

$$\left|\arg F_w(z) - \tan^{-1}\frac{\mathrm{Im}(A\overline{B})r^2}{\mathrm{Re}(A\overline{B})r^2 - 1}\right| < \sin^{-1}\frac{|A-B|r}{|1-A\overline{B}r^2|}.$$

Since $\log f(z) = \log|f(z)| + i\arg f(z)$ and $p(z) = \dfrac{1}{A-B}\log\left(F_w(z)\right)$, thus for $A = -B = \alpha$ we obtain (2.2) and (2.3) and for $A = \alpha e^{i\gamma}$ and $B = \alpha e^{-i\gamma}$, we have (2.4) and (2.5). ∎

For the sake of computational convenience, we shall make the following assumptions:

$$h_1 := \frac{1}{2\alpha}\log\left(\frac{1+\alpha}{1-\alpha}\right), \qquad h_2 := \frac{1}{2\alpha}\sin^{-1}\left(\frac{2\alpha}{1+\alpha^2}\right),$$

$$k_1 := \frac{1}{2\alpha\sin\gamma}\sin^{-1}\left(\frac{2\alpha\sin\gamma}{\sqrt{1+\alpha^4-2\alpha^2\cos 2\gamma}}\right),$$

$$k_2 := \frac{1}{2\alpha\sin\gamma}\log\frac{\sqrt{1+\alpha^4-2\alpha^2\cos 2\gamma}+2\alpha\sin\gamma}{1-\alpha^2}$$

and

$$k := \frac{1}{2\alpha\sin\gamma}\tan^{-1}\left(\frac{\alpha^2\sin 2\gamma}{\alpha^2\cos 2\gamma - 1}\right).$$

Now we shall discuss the image domain of $\psi_{A,B}$ for different cases of α:
For $0 < \alpha < 1$, $\psi_{A,B}(\mathbb{D}) = \{u+iv \in \mathbb{C} : (u,v) \in \Omega_1\}$ where

$$\Omega_1 := \begin{cases} \dfrac{u^2}{h_1^2} + \dfrac{v^2}{h_2^2} < 1, & \text{when } A = -B = \alpha \\ \dfrac{(u-k)^2}{k_1^2} + \dfrac{v^2}{k_2^2} < 1, & \text{when } A = \alpha e^{i\gamma}, B = \alpha e^{-i\gamma} \text{ for } 0 < \gamma \leq \pi/2. \end{cases} \tag{2.6}$$

Moreover when $\alpha = 1$, the major axis h_1 and k_2 respectively, tends to infinity, hence $\psi_{A,B}(\mathbb{D}) = \{u+iv \in \mathbb{C} : (u,v) \in \Omega_2\}$ where Ω_2 is the following strip domain

$$\Omega_2 := \begin{cases} -\dfrac{\pi}{4} < v < \dfrac{\pi}{4}, & \text{when } A = -B = 1 \\ \dfrac{\gamma-\pi}{2\sin\gamma} < u < \dfrac{\gamma}{2\sin\gamma}, & \text{when } A = e^{i\gamma}, B = e^{-i\gamma} \text{ for } 0 < \gamma \leq \pi/2. \end{cases} \tag{2.7}$$

Remark 2.4. (i) $\max_{|z|=r}\mathrm{Re}\,\psi_{A,B}(z) = \psi_{A,B}(r)$ and $\min_{|z|=r}\mathrm{Re}\,\psi_{A,B}(z) = \psi_{A,B}(-r)$.

(ii) If $\alpha < 1$, then $p \prec \psi_{A,B}(z)$ if and only if $p(\mathbb{D}) \subseteq \Omega_1$ and if $\alpha = 1$, then $p \prec \psi_{A,B}(z)$ if and only if $p(\mathbb{D}) \subsetneq \Omega_2$, where Ω_1 and Ω_2 are as defined in (2.6) and (2.7) respectively.

(iii) $1 + \psi_{A,B} \in \widetilde{\mathcal{M}}_A$.

Now in the following lemma, we aim to find the radius of the smallest (largest) disk centered at the point (1, 0) that can subscribe (inscribe) the domain $1+\psi_{A,B}(\mathbb{D})$.

Lemma 2.5. *Let $\alpha \in (0,1)$ and $\gamma \in (0, \pi/2]$, then we have*

(i) for $A = -B = \alpha$,

$$\{w \in \mathbb{C} : |w-1| < h_2\} \subset 1+\psi_{A,B}(\mathbb{D}) \subset \{w \in \mathbb{C} : |w-1| < h_1\}.$$

(ii) for $A = \alpha e^{i\gamma}$ and $B = \alpha e^{-i\gamma}$,

$$\{w \in \mathbb{C} : |w-1| < k_1 + k\} \subset 1+\psi_{A,B}(\mathbb{D}) \subset \{w \in \mathbb{C} : |w-1| < k_2\}.$$

Further for $\alpha = 1$, we have

(iii) for $A = -B = 1$,

$$\left\{w \in \mathbb{C} : |w-1| < \frac{\pi}{4}\right\} \subset 1+\psi_{A,B}(\mathbb{D}).$$

(iv) for $A = e^{i\gamma}$ and $B = e^{-i\gamma}$,

$$\left\{w \in \mathbb{C} : |w-1| < \frac{\gamma}{2\sin\gamma}\right\} \subset 1+\psi_{A,B}(\mathbb{D}).$$

Proof. Let $w = 1+\psi_{A,B}(z)$, then we have

$$|w-1| = |\psi_{A,B}(z)| = \frac{1}{|A-B|}\left|\log\left(\frac{1+Az}{1+Bz}\right)\right|.$$

It can be easily seen that for $\alpha \in (0,1)$, $A = -B = \alpha$

$$\min_{|z|=1}|\psi_{A,B}(z)| = h_2 \text{ and } \max_{|z|=1}|\psi_{A,B}(z)| = h_1$$

and for $A = \alpha e^{i\gamma}$ and $B = \alpha e^{-i\gamma}$

$$\min_{|z|=1}|\psi_{A,B}(z)| = k_1 + k \text{ and } \max_{|z|=1}|\psi_{A,B}(z)| = k_2.$$

Further for $\alpha = 1$, h_1 and k_2 tends to ∞, $h_2 = \pi/4$ and $k_1 + k = \gamma/(2\sin\gamma)$. This completes the proof. ∎

Remark 2.6. In general, if we consider a disc $\mathcal{D}(a,r) := \{w \in \mathbb{C}; |w-a| < r\}$ where $a \in \mathbb{R}$ and $1 \in \mathcal{D}(a,r)$, then we observe that $\mathcal{D}(a,r) \subset 1+\psi_{A,B}(\mathbb{D})$ if and only if

$$\mathcal{D}(a,r) \subset \begin{cases} \mathcal{D}(1,h_2), & \text{for } A = -B = \alpha, \\ \mathcal{D}(1+k,k_1), & \text{for } A = \alpha e^{i\gamma},\ B = \alpha e^{-i\gamma}. \end{cases} \tag{2.8}$$

Thus we arrive at the following:

Lemma 2.7. *Let $-1 < D < C \le 1$ and $p(z) := (1+Cz)/(1+Dz)$. Then $p(z) \prec \mathcal{L}[A,B]$ if and only if*

(i) for $A = -B = \alpha$

$$C \le \begin{cases} h_2 + (1-h_2)D, & \text{when } (1-CD)/(1-D^2) \le 1, \\ h_2 + (1+h_2)D, & \text{when } (1-CD)/(1-D^2) \ge 1. \end{cases} \tag{2.9}$$

(ii) for $A = \alpha e^{i\gamma}$ and $B = \alpha e^{-i\gamma}$,

$$C \le \begin{cases} k_1 - k + (1-k_1+k)D, & \text{when } (1-CD)/(1-D^2) \le 1+k, \\ k_1 + k + (1+k_1+k)D, & \text{when } (1-CD)/(1-D^2) \ge 1+k. \end{cases} \tag{2.10}$$

Proof. Since $p(z) = (1+Cz)/(1+Dz)$ maps $\mathbb{D}$ onto $\mathcal{D}(a,r)$, where

$$a = \frac{1-CD}{1-D^2} \quad \text{and} \quad r = \frac{C-D}{1-D^2},$$

the result follows at once from (2.8). ∎

From Lemma 2.7, we deduce the following corollary, which ensures that $\mathcal{F}[A,B]$ is nonempty.

Corollary 2.8. *Let $-1 < D < C \leq 1$ and $f \in \mathcal{A}$ be such that $zf'(z)/f(z) = (1+Cz)/(1+Dz)$, then $f \in \mathcal{F}[A,B]$ if and only if, the condition (2.9) or (2.10) holds.*

Now from (1.4), we have $f \in \mathcal{F}[A,B]$ if and only if there exists an analytic function $p(z) \prec \psi_{A,B}(z)$ such that

$$f(z) = z \exp\left(\int_0^z \frac{p(t)}{t} dt\right). \tag{2.11}$$

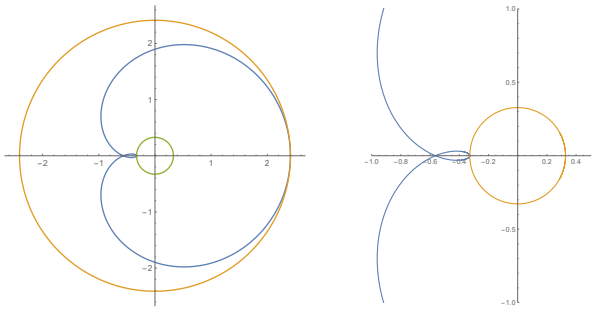
If we take $p(z) = \psi_{A,B}(z)$, then we obtain from (2.11)

$$f_{A,B}(z) := z \exp \frac{Li_2(-Bz) - Li_2(-Az)}{A-B}, \tag{2.12}$$

where $Li_2(x) = \sum_{n=1}^{\infty} x^n/n^2$, denotes the Spence's (or dilogarithm) function. The function $f_{A,B}(z)$ is nonunivalent but have the extremal properties for many problems for the class $\mathcal{F}[A,B]$ (see Fig:1). Kumar and Gangania [5] studied the class $\mathcal{F}(\psi)$ as defined in (1.2) and obtained growth and covering theorem for the case when $1+\psi(z) \not\prec (1+z)/(1-z)$. As a consequence of the same and Remark 2.4, we obtain the following result:

Lemma 2.9. *Let $f \in \mathcal{F}[A,B]$ and $f_{A,B}$ be defined as in (2.12), then for $|z| = r$*

(i) Growth Theorem: $-f_{A,B}(-r) \leq |f(z)| \leq f_{A,B}(r)$.
(ii) Covering Theorem: Either f is a rotation of $f_{A,B}$ or $\{w \in \mathbb{C} : |w| \leq -f_{A,B}(-1)\} \subset f(\mathbb{D})$.



(A) (B)

FIGURE 1. (A) The images of $\partial\mathbb{D}_{f_{A,B}(1)}$, $f_{A,B}(\partial\mathbb{D})$ and $\partial\mathbb{D}_{f_{A,B}(-1)}$, for $A = 0.5e^{i\pi/3}$ and $B = 0.5e^{-i\pi/3}$ and (B) zoomed image of cusp.

For $z = re^{i\theta}$, where θ is fixed but arbitrary, as a consequence of growth theorem and $\psi_{A,B}(-r) \leq \operatorname{Re} \psi_{A,B}(re^{i\theta}) \leq \psi_{A,B}(r)$, we obtain

$$\log \frac{f(z)}{z} = \int_0^r \frac{p(te^{i\theta})}{t} dt,$$

where $p(z) := \psi_{A,B}(w(z))$ and w is a Schwarz functon. Further

$$\frac{f(z)}{z} = \exp\int_0^r \frac{p(te^{i\theta})}{t}dt = \exp\left(\int_0^r \operatorname{Re}\frac{p(te^{i\theta})}{t}dt + i\int_0^r \operatorname{Im}\frac{p(te^{i\theta})}{t}dt\right)$$

$$\left|\frac{f(z)}{z}\right| \leq \exp\int_0^r \frac{\operatorname{Re}\psi_{A,B}(te^{i\theta})}{t}dt$$

$$\exp\int_0^r \frac{\psi_{A,B}(-t)}{t}dt \leq \left|\frac{f(z)}{z}\right| \leq \exp\int_0^r \frac{\psi_{A,B}(t)}{t}dt.$$

Here $L(f,r) := \int_0^{2\pi}|zf'(z)|d\theta$ is the length of the boundary curve $f(|z|= r)$. Now we obtain the following result:

Corollary 2.10. *Let $f \in \mathcal{F}[A,B]$ and $M(r) = \exp\int_0^r \frac{\psi_{A,B}(t)}{t}dt$, then for $|z|= r$, we have*

(i) $M(-r) \leq \left|\frac{f(z)}{z}\right| \leq M(r)$.
(ii) $(1+\max_{|z|\leq r}|\psi_{A,B}(z)|)M(-r) \leq |f'(z)| \leq (1+\max_{|z|\leq r}|\psi_{A,B}(z)|)M(r)$.
(iii) $2\pi r(1+\max_{|z|\leq r}|\psi_{A,B}(z)|)M(-r) \leq L(f,r) \leq 2\pi r(1+\max_{|z|\leq r}|\psi_{A,B}(z)|)M(r)$.
(iv) $f(z)/z \prec f_{A,B}(z)/z$.

3. Radius Estimates

In this section, we obtain some radii estimates. From Theorem 2.3 we conclude that $f \in \mathcal{S}[A,B]$ is starlike of order $1+\frac{1}{|A-B|}(\varphi-\vartheta) < 0$, hence f may not be univalent in $\mathbb{D}$, thus $\mathcal{F}[A,B] \not\subseteq \mathcal{S}^*$. Therefore, in the following result, we find the radius of starlikeness of order δ, where $\delta \in [0,1)$ for functions in the class $\mathcal{F}[A,B]$.

Theorem 3.1. *Let $\alpha \in (0,1]$, $\gamma \in (0,\gamma_0)$, where $\gamma_0 \simeq 1.2461\ldots$ and $\delta \in [0,1)$ be given numbers. If $f \in \mathcal{F}[A,B]$, then f is starlike of order δ in the disc $|z|< r(\delta)$, where*

(i) For $A = -B = \alpha$,

$$r(\delta) = \frac{\exp(2\alpha(1-\delta))-1}{\alpha(\exp(2\alpha(1-\delta))+1)}. \tag{3.1}$$

(ii) For $A = \alpha e^{i\gamma}$ and $B = \alpha e^{-i\gamma}$, $r(\delta)$ is the smallest positive root of

$$\tan(2\alpha\sin\gamma(\delta-1)) = \frac{\alpha^4r^4\sin 2\gamma + 2\alpha^3r^3\sin\gamma\cos 2\gamma - \alpha^2r^2\sin 2\gamma - 2\alpha r\sin\gamma}{\alpha^4r^4\cos 2\gamma - 2\alpha^3r^3\sin\gamma\sin 2\gamma - \alpha^2r^2\cos 2\gamma - \alpha^2r^2 + 1}. \tag{3.2}$$

The result is sharp.

Proof. Let $f \in \mathcal{F}[A,B]$, then using Theorem 2.3 for $|z|< r$, we have
Case 1. when $A = -B = \alpha$,

$$\operatorname{Re}\frac{zf'(z)}{f(z)} > 1 - \frac{1}{2\alpha}\log\left(\frac{1+\alpha r}{1-\alpha r}\right) =: t_1(r),$$

Now as $t_1'(r) < 0$ for all choices of α, thus $t_1(r)$ is strictly decreasing function from 1 to $1 - 1/(2\alpha)\log((1+\alpha)/(1-\alpha)) < 0$. Hence the root $r(\delta)$, given in (3.1), of the equation $t_1(r) = \delta$, is the radius of starlikeness of order δ of $\mathcal{F}[A,B]$.
Case 2. when $A = \alpha e^{i\gamma}$ and $B = \alpha e^{-i\gamma}$

$$\operatorname{Re}\frac{zf'(z)}{f(z)} > 1 - \frac{1}{2\alpha\sin\gamma}(\eta-\tau) =: t_2(r),$$

where η and τ are given in Theorem 2.3. Since $t_2(r) = \delta$ can be simplified to the equation (3.2) by a simple computation, thus it is enough to find the smallest positive root of (3.2). Further

$$t_2'(r) = -\frac{1-\alpha^4 r^4 + 2\alpha r\cos\gamma(1-\alpha^2 r^2)}{(1-\alpha^2 r^2)\sqrt{1+\alpha^4 r^4 - 2\alpha^2 r^2\cos 2\gamma}} < 0$$

for all choices of α and γ, $t_2(r)$ is a strictly decreasing function from 1 to $g(\alpha,\gamma)$, where

$$g(\alpha,\gamma) := 1 + \frac{1}{2\alpha\sin\gamma}\left(\tan^{-1}\left(\frac{\alpha^2\sin 2\gamma}{\alpha^2\cos 2\gamma - 1}\right) - \sin^{-1}\left(\frac{2\alpha\sin\gamma}{\sqrt{1+\alpha^4-2\alpha^2\cos 2\gamma}}\right)\right).$$

Note that $g(\alpha,\gamma)$ is real valued, non-constant analytic function on a bounded domain $R := (0,1)\times$

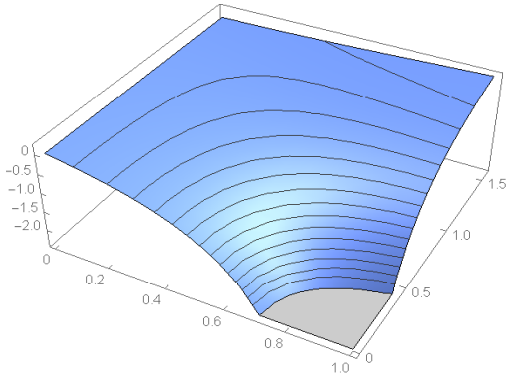


FIGURE 2. Image of $g(\alpha,\gamma)$ on $R := (0,1)\times(0,\pi/2)$.

$(0,\pi/2)$ and is bounded above on R. Thus by Maximum Modulus Theorem, $g(\alpha,\gamma)$ attains its maximum value on the boundary of R and the four boundaries are enlisted below:

1. $g(0,\gamma) = 0$.
2. $g(1,\gamma) = 1 + (\gamma-\pi)/(2\sin\gamma)$. Now as
$$g'(1,\gamma) = (1+(\pi-\gamma)\cot\gamma)/(2\sin\gamma) > 0,$$
therefore $g(1,\gamma)$ is strictly increasing from $-\infty$ to $1-\pi/4 \simeq 0.2146\ldots$ and γ_0 is the zero of it.
3. $g(\alpha,0) = 1-1/(1-\alpha)$. As $g'(\alpha,0) = \log(1-\alpha) < 0$, therefore $g(\alpha,0)$ is strictly decreasing from 0 to $-\infty$.
4. $g(\alpha,\pi/2) = 1 - \left(\sin^{-1}\left(2\alpha/(1+\alpha^2)\right)\right)/(2\alpha)$. As
$$g'(\alpha,\pi/2) = \left(\left(2\alpha/(1+\alpha^2)\right) + \sin^{-1}\left(2\alpha/(1+\alpha^2)\right)\right)/(2\alpha^2) > 0,$$
therefore $g(\alpha,\pi/2)$ is also strictly increasing from 0 to $1-\pi/4 \simeq 0.2146$.

From Fig. 2 and all four boundaries, we observe that $g(\alpha,\gamma) < 0$ for all choices of $\alpha \in (0,1]$, $\gamma \in (0,\gamma_0)$, where γ_0 is the root of the equation $g(1,\gamma) = 0$. Hence $r(\delta)$ is the radius of starlikeness of order δ of $\mathcal{F}[A,B]$. ∎

Theorem 3.2. *Let $\alpha \in (0,1]$ and $\delta \in [0,1)$. If $f \in \mathcal{F}[A,B]$, for the case when $A = \alpha e^{i\gamma}$ and $B = \alpha e^{-i\gamma}$, then there is an unique $\gamma' \in (\gamma_0,\pi/2)$ such that*

(i) when $\gamma \in (\gamma_0,\gamma']$, f is starlike of order δ in $\mathbb{D}_{r(\delta)}$, where $r(\delta)$ the root of the equation (3.2).
(ii) when $\gamma \in (\gamma',\pi/2)$, f is starlike of order $\delta \in [0, g(\alpha,\pi/2))$ in $\mathbb{D}$ and starlike of order $\delta \in [g(\alpha,\pi/2),1)$ in $\mathbb{D}_{r(\delta)}$.

Proof. By observing the Fig. 2 and the nature of all boundary curves discussed in Theorem 3.1, we arrive at $g(\alpha,\gamma_0)<0$ and $g(\alpha,\pi/2)>0$ for all $\alpha\in(0,1)$. Thus by IVP for any choice of $\alpha\in(0,1)$, there is an unique $\gamma'\in(\gamma_0,\pi/2)$ such that $g(\alpha,\gamma')=0$, $g(\alpha,\gamma)<0$ for all $\gamma\in(\gamma_0,\gamma')$ and $g(\alpha,\gamma)>0$ for all $\gamma\in(\gamma',\pi/2)$. Hence the result follows. ∎

Taking $\delta=0$ in Theorems 3.1 and 3.2, we obtain the following result:

Corollary 3.3. *Let $\alpha\in(0,1]$ and $\gamma\in(0,\pi/2]$. If $f\in\mathcal{F}[A,B]$ then*

(i) For $A=-B=\alpha$, f is starlike univalent in the disc $|z|<r_0$, where

$$r_0=\frac{\exp(2\alpha)-1}{\alpha(\exp(2\alpha)+1)}.$$

(ii) For $A=\alpha e^{i\gamma}$ and $B=\alpha e^{-i\gamma}$, $f\in\mathcal{S}^*$, whenever $\gamma\in(\gamma',\pi/2)$, and f is starlike univalent in the disc $|z|<r_0$, whenever $\gamma\in(0,\gamma_0)\cup(\gamma_0,\gamma']$, where γ_0, γ' are given in Theorems 3.1 and 3.2, respectively and r_0 is the smallest positive root of

$$\tan(2\alpha\sin\gamma)+\frac{\alpha^4r^4\sin2\gamma+2\alpha^3r^3\sin\gamma\cos2\gamma-\alpha^2r^2\sin2\gamma-2\alpha r\sin\gamma}{\alpha^4r^4\cos2\gamma-2\alpha^3r^3\sin\gamma\sin2\gamma-\alpha^2r^2\cos2\gamma-\alpha^2r^2+1}=0.$$

The result is sharp.

Corollary 3.4. *Let $f\in\mathcal{F}[A,B]$. Then for the disc $|z|<r\le r_0$, where r_0 is given in Corollary 3.3, we have*

(i) For $A=-B=\alpha$,

$$\left|\arg\left(\frac{zf'(z)}{f(z)}\right)\right|<\tan^{-1}\left(\frac{\sin^{-1}\left(\frac{2\alpha r}{1+\alpha^2r^2}\right)}{2\alpha-\log\left(\frac{1+\alpha r}{1-\alpha r}\right)}\right)$$

(ii) For $A=\alpha e^{i\gamma}$ and $B=\alpha e^{-i\gamma}$,

$$\left|\arg\left(\frac{zf'(z)}{f(z)}\right)\right|<\tan^{-1}\left(\frac{\log T_2}{2\alpha\sin\gamma-\eta+\tau}\right),$$

where T_2, η and τ are given in Theorem 2.3.

Corollary 3.5. *Let $f\in\mathcal{F}[A,B]$. Then for the disc $|z|<r_s\le r_0$, where r_0 is given in Corollary 3.3, $f\in\mathcal{SS}^*(\beta)$, where $r_s\in(0,r_0]$ is the smallest positive root of the following equation:*

(i) For $A=-B=\alpha$,

$$\sin^{-1}\left(\frac{2\alpha r}{1+\alpha^2r^2}\right)=\tan\left(\frac{\beta\pi}{2}\right)\left(2\alpha-\log\left(\frac{1+\alpha r}{1-\alpha r}\right)\right)$$

(ii) For $A=\alpha e^{i\gamma}$ and $B=\alpha e^{-i\gamma}$,

$$\frac{\log T_2}{2\alpha\sin\gamma-\eta+\tau}=\tan\left(\frac{\beta\pi}{2}\right),$$

where T_2, η and τ are given in Theorem 2.3.

We now discuss the radius estimates for the classes $\mathcal{BS}^*(\alpha)$ and $\mathcal{S}^*_{cs}$, which are not contained in $\mathcal{S}^*$. Further we derive radii for functions in $\mathcal{BS}^*(\alpha)$ and $\mathcal{S}^*_{cs}$ to be in $\mathcal{F}[A,B]$.

Theorem 3.6. *Let $\alpha\in(0,1)$, $r_0=(-1+\sqrt{1+4\alpha h_1^2})/(2\alpha h_1)$ and $\alpha_0\in(0,1)$ is the smallest solution of*

$$1-\frac{r_0^2\cos^2\theta(1-\alpha r_0^2)^2}{(1+\alpha^2r_0^4-2\alpha r_0^2\cos2\theta)^2h_1^2}-\frac{r_0^2\sin^2\theta(1+\alpha r_0^2)^2}{(1+\alpha^2r_0^4-2\alpha r_0^2\cos2\theta)^2h_2^2}=0,$$

where $\theta \in (0, \pi/2)$. If $f \in \mathcal{BS}^*(\alpha)$ then for the disc $|z| \leq r_b < 1$, $f \in \mathcal{F}[A, B]$, where

$$r_b = \begin{cases} r_0, & \text{for } A = -B = \alpha \leq \alpha_0 \\ r_1 := \dfrac{-1 + \sqrt{1 + 4\alpha h_2^2}}{2\alpha h_2}, & \text{for } A = -B = \alpha > \alpha_0 \\ r_2 := \dfrac{-1 + \sqrt{1 + 4\alpha(k_1 + k)^2}}{2\alpha(k_1 + k)}, & \text{for } A = \alpha e^{i\gamma} \text{ and } B = \alpha e^{-i\gamma}. \end{cases}$$

Moreover the radii r_0 and r_2 are sharp.

Proof. Let $f \in \mathcal{BS}^*(\alpha)$. Since for $A = -B = \alpha$, $\max_{|z|=1} \operatorname{Re} \psi_{A,B}(z) = h_1$. Thus to find such $r < 1$ for which the image of $zf'(z)/f(z) - 1$ under the disc $|z| < r$ lies inside $\psi_{A,B}(\mathbb{D})$, it is necessary that

$$\max_{|z|=r<1} \operatorname{Re}\left(\frac{z}{1 - \alpha z^2}\right) = \frac{r}{1 - \alpha r^2} \leq h_1 \tag{3.3}$$

must hold. Clearly for $|z| \leq r_0$, the inequality (3.3) holds. Now to see that for $\alpha \leq \alpha_0$, radius r_0 is also sufficient for $zf'(z)/f(z) - 1 \prec z/(1 - \alpha z^2) \in \psi_{A,B}(\mathbb{D})$ in the disc $|z| \leq r_0$. For $\zeta = re^{i\theta}$ ($\theta \in [0, 2\pi)$), we have

$$B_r(\theta) := \frac{\zeta}{1 - \alpha\zeta^2} = \frac{r\cos\theta(1 - \alpha r^2)}{1 + \alpha^2 r^4 - 2\alpha r^2 \cos 2\theta} + i\frac{r\sin\theta(1 + \alpha r^2)}{1 + \alpha^2 r^4 - 2\alpha r^2 \cos 2\theta}.$$

Since $\operatorname{Re} B_r(\theta) = \operatorname{Re} B_r(-\theta)$, $\operatorname{Re} B_r(\theta) = -\operatorname{Re} B_r(\pi - \theta)$ and $\operatorname{Im} B_r(\theta) = \operatorname{Im} B_r(\pi - \theta)$, therefore the curve $B_r(\theta)$ is symmetric about real and imaginary axis thus it is sufficient to consider for $\theta \in [0, \pi/2]$. Now for $r = r_0$, the square of the distance from the origin to the points of $B_{r_0}(\theta)$ is given by

$$Dist(0; B_{r_0}(\theta)) := \frac{r_0^2}{1 + \alpha^2 r_0^4 - 2\alpha r_0^2 \cos 2\theta}.$$

Since $Dist(0; B_{r_0}(\theta))' < 0$, thus $Dist(0; B_{r_0}(\theta))$ is a decreasing function of θ. Hence the farthest point of $B_{r_0}(\theta)$ from origin is $(r_0/(1 - \alpha r_0^2), 0)$, which lies on the boundary of Ω_1. Now for $A = -B = \alpha > \alpha_0$ and $A = \alpha e^{i\gamma}$, $B = \alpha e^{-i\gamma}$ with $|z| = r$, we have

$$\left|\frac{zf'(z)}{f(z)} - 1\right| \leq \frac{r}{1 - \alpha r^2}.$$

Therefore, by Lemma 2.5, we see that $\mathcal{F}[A, B]-$radius for the class $\mathcal{BS}^*(\alpha)$ are the smallest positive roots r_1 and r_2 of the equations $r/(1 - \alpha r^2) = h_2$ and $r/(1 - \alpha r^2) = k_1 + k$, respectively. ∎

Theorem 3.7. *Let $\alpha \in (0, 1)$ and $\alpha_0 \in (0, 1)$ be the smallest solution of*

$$\frac{r_0^2((\alpha - 1)r_0 + \cos\theta(1 - \alpha r_0^2))^2}{h_1^2} + \frac{r_0^2(1 + \alpha r_0^2)^2 \sin^2\theta}{h_2^2} = N,$$

where $\theta \in (0, \pi/2)$ and

$$N = (1 + r_0^2 - 2r_0\cos\theta)^2(1 + \alpha^2 r_0^2 + 2\alpha r_0 \cos\theta)^2$$

and

$$r_0 = (-(1 + (1 - \alpha)h_1) + \sqrt{(1 + (1 - \alpha)h_1)^2 + 4\alpha h_1^2})/(2\alpha h_1).$$

If $f \in \mathcal{S}_{cs}^*(\alpha)$ then for the disc $|z| \leq r_{cs} < 1$, $f \in \mathcal{F}[A, B]$, where

$$r_{cs} = \begin{cases} r_0, & \text{for } A = -B = \alpha \leq \alpha_0 \\ r_1 := \dfrac{-(1 + (1 - \alpha)h_2) + \sqrt{(1 + (1 - \alpha)h_2)^2 + 4\alpha h_2^2}}{2\alpha h_2}, & \text{for } A = -B = \alpha > \alpha_0 \\ r_2 := \dfrac{-(1 + (1 - \alpha)(k_1 + k)) + \sqrt{(1 + (1 - \alpha)(k_1 + k))^2 + 4\alpha(k_1 + k)^2}}{2\alpha(k_1 + k)}, & \text{for } A = \alpha e^{i\gamma}, B = \alpha e^{-i\gamma}. \end{cases}$$

Moreover the radii r_0 and r_2 are sharp.

Proof. Let $f \in \mathcal{S}^*_{cs}(\alpha)$. Since for $A = -B = \alpha$, $\max_{|z|=1} \operatorname{Re} \psi_{A,B}(z) = h_1$. Thus to find such $r < 1$ for which the image of $zf'(z)/f(z) - 1$ under the disc $|z| < r$ lies inside $\psi_{A,B}(\mathbb{D})$, it is necessary that

$$\max_{|z|=r<1} \operatorname{Re}\left(\frac{z}{(1-z)(1+\alpha z)}\right) = \frac{r}{(1-r)(1+\alpha r)} \leq h_1 \tag{3.4}$$

must hold. Clearly for $|z| \leq r_0$, the equation (3.4) holds. Now to see that for $\alpha \leq \alpha_0$ radius r_0 is also sufficient for $zf'(z)/f(z) - 1 \prec z/((1-z)(1+\alpha z)) \in \psi_{A,B}(\mathbb{D})$ in the disc $|z| \leq r_0$. For $\zeta = re^{i\theta}$ ($\theta \in [0, 2\pi)$), we have

$$CS_r(\theta) := \frac{\zeta}{(1-\zeta)(1+\alpha\zeta)} = \frac{r((\alpha-1)r + \cos\theta(1-\alpha r_0^2))}{(1+r^2-2r\cos\theta)(1+\alpha^2 r^2+2\alpha r\cos\theta)} + i\frac{r(1+\alpha r_0^2)\sin\theta}{(1+r^2-2r\cos\theta)(1+\alpha^2 r^2+2\alpha r\cos\theta)}.$$

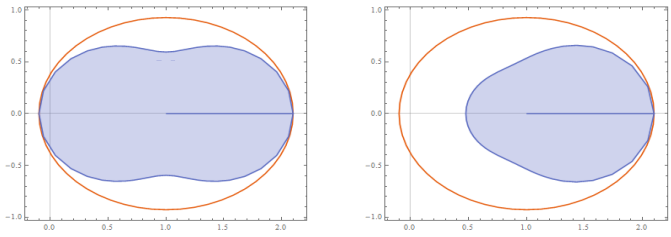
Since $\operatorname{Re} CS_r(\theta) = \operatorname{Re} CS_r(-\theta)$, therefore the curve $CS_r(\theta)$ is symmetric about real axis thus it is sufficient to consider for $\theta \in [0, \pi]$. Now for $r = r_0$, the square of the distance from the origin to the points of $CS_{r_0}(\theta)$ is given by

$$Dist(0; CS_{r_0}(\theta)) := \frac{r_0^2}{(1+r^2-2r\cos\theta)^2(1+\alpha^2 r^2+2\alpha r\cos\theta)^2}.$$

Since $Dist(0; CS_{r_0}(\theta))' = 0$ for $\theta = 0, \theta_0$ and π with $Dist(0; CS_{r_0}(\theta))' < 0$ whenever $\theta \in (0, \theta_0)$ and $Dist(0; CS_{r_0}(\theta))' > 0$ whenever $\theta \in (\theta_0, \pi)$. And $Dist(0; CS_{r_0}(0)) - Dist(0; CS_{r_0}(\pi)) > 0$, hence the farthest point of $CS_{r_0}(\theta)$ from origin is equal to h_1 obtained at $\theta = 0$. Now for $A = -B = \alpha > \alpha_0$ and $A = \alpha e^{i\gamma}$, $B = \alpha e^{-i\gamma}$ with $|z| = r$, we have

$$\left|\frac{zf'(z)}{f(z)} - 1\right| \leq \frac{r}{(1-r)(1+\alpha r)}.$$

Therefore, by Lemma 2.5, we see that $\mathcal{F}[A, B]-$radius for the class $\mathcal{S}^*_{cs}(\alpha)$ are the smallest positive roots r_1 and r_2 of the equations $r/((1-r)(1+\alpha r)) = h_2$ and $r/((1-r)(1+\alpha r)) = k_1 + k$, respectively. ∎



(A) (B)

FIGURE 3. $f_\alpha(z) = z/(1-\alpha z^2)$ and $g_\alpha(z) = z/((1-z)(1+\alpha z))$, (A) $f_{0.5}(\mathbb{D}_{r_0=0.77}) \subset \psi_{0.5,-0.5}(\partial\mathbb{D})$ and (B) $g_{0.5}(\mathbb{D}_{r_0=0.59}) \subset \psi_{0.5,-0.5}(\partial\mathbb{D})$.

References

[1] P. Goel and S. Sivaprasad Kumar, Certain class of starlike functions associated with modified sigmoid function, Bull. Malays. Math. Sci. Soc. **43** (2020), no. 1, 957–991.
[2] R. Kargar, A. Ebadian and J. Sokół, On Booth lemniscate and starlike functions, Anal. Math. Phys. **9** (2019), no. 1, 143–154.
[3] S. S. Kumar and S. Banga, On a Special Type of Ma-Minda Function. arxiv preprint arXiv:2006.02111, (2020).
[4] S. S. Kumar and P. Yadav, On Oblique Domains of Janowski Function. arXiv preprint arXiv:2104.08533 (2021).
[5] S. S. Kumar and G. Kamaljeet, On Geometrical Properties of Certain Analytic functions, Iran. J. Sci. Technol. Trans. A Sci. (2020), https://doi.org/10.1007/s40995-021-01116-1.
[6] W. C. Ma and D. Minda, A unified treatment of some special classes of univalent functions, in *Proceedings of the Conference on Complex Analysis (Tianjin, 1992)*, 157–169, Conf. Proc. Lecture Notes Anal., I, Int. Press, Cambridge.
[7] V. S. Masih, A. Ebadian and S. Yalçin, Some properties associated to a certain class of starlike functions, Math. Slovaca **69** (2019), no. 6, 1329–1340.
[8] V. S. Masih and S. Kanas, Subclasses of Starlike and Convex Functions Associated with the Limaçon Domain, Symmetry **12** (2020), no. 6, 942.
[9] R. Mendiratta, S. Nagpal and V. Ravichandran, On a subclass of strongly starlike functions associated with exponential function, Bull. Malays. Math. Sci. Soc. **38** (2015), no. 1, 365–386.
[10] R. K. Raina and J. Sokół, Some properties related to a certain class of starlike functions, C. R. Math. Acad. Sci. Paris **353** (2015), no. 11, 973–978.
[11] M. S. Robertson, Certain classes of starlike functions, Michigan Math. J. **32** (1985), no. 2, 135–140.
[12] K. Sakaguchi, A note on p-valent functions, J. Math. Soc. Japan **14** (1962), 312–321.
[13] J. Sokół and J. Stankiewicz, Radius of convexity of some subclasses of strongly starlike functions, Zeszyty Nauk. Politech. Rzeszowskiej Mat. No. 19 (1996), 101–105.
[14] B. A. Uralegaddi, M. D. Ganigi and S. M. Sarangi, Univalent functions with positive coefficients, Tamkang J. Math. **25** (1994), no. 3, 225–230.

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
Email address: spkumar@dce.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
Email address: poojayv100@gmail.com

View publication stats

Electrical Engineering
https://doi.org/10.1007/s00202-023-01945-w

ORIGINAL PAPER

Optimal design of renewable energy based hybrid system considering weather forecasting using machine learning techniques

Bandana Sharma[1] · **M. Rizwan**[1] · **P. Anand**[2]

Received: 1 April 2023 / Accepted: 11 July 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract
Renewable energy resources are becoming more appealing as energy demand and fossil fuel costs increase. The hybridization of these resources has the potential to reduce unpredictability and intermittency while increasing efficiency. The accuracy of size optimization of hybrid renewable energy systems (HRES) can be improved by using accurate weather data that can be obtained through forecasting. Thus, to increase the precision of the size optimization, hourly forecasting of global horizontal irradiation, temperature, and wind speed for one year has been performed using Gaussian process regression (GPR), Support Vector Regression, Extreme Gradient Boosting, and Decision Tree techniques. The results of all four forecasting models (FM) are then compared and revealed that the results obtained from GPR are better than those of other FM; therefore, the forecasted data for solar, wind, and temperature obtained from GPR are used for sizing the HRES. The net present cost is utilized to analyze the viability of the HRES while considering system reliability. Furthermore, recently developed optimization algorithms, namely the Colony Predation Algorithm (CPA), Tunicate Swarm Algorithm (TSA), and Aquila Optimization (AO) algorithms have been applied to the sizing of a grid-connected HRES to meet the energy needs of a remote site in the Indian province of Haryana. A comparison of CPA, AO, and TSA has been carried out and revealed that TSA offers more promising outcomes. In addition, the simulation results demonstrate a 0.42% reduction in per unit cost of energy when forecasted data has been used for size optimization.

Keywords Hybrid renewable energy system (HRES) · Optimization techniques · Forecasting · Machine Learning · Rural electrification

1 Introduction

Globally, the focus has been switched from conventional energy to green energy due to the escalating pollution rates and catastrophic repercussions of climate change that endanger life on the earth. However, ubiquitous and environmentally sustainable renewable energy sources (RES) are unpredictable and location-dependent. Now, it is recognized that only HRES with blended generation and adequate storage can meet the need for a consistent, ecological, and reliable power supply; as a result, the optimal sizing of such systems has turned into a prominent topic for researchers around the world [1]. Over the preceding two decades, various comprehensive techniques have been employed for developing and determining the optimal size of HRESs. These techniques incorporate simulation tools, graphical construction [2], probabilistic [3], iterative [4, 5], and artificial intelligence [6–11]. In recent years, intelligent techniques have been widely used to optimize HRES and boost economic rewards [12]. Several researchers have used intelligent approaches for the optimal sizing of HRES; some are discussed in the following paragraphs.

Bernal-Agustn et al. and Dufo-López et al. researched the optimal size of an HRES of solar photovoltaic (SPV)/wind turbine (WT)/diesel generator. In this work, the Strength Pareto evolutionary algorithm involving two target functions: system cost reduction and greenhouse gas emission reduction, has been applied [13, 14]. Yang et al., Bilal et al., and Tafreshi et al. applied the genetic algorithm (GA) to optimize the sizing problem under a variety of objective functions, including maximizing the system reliability under

✉ Bandana Sharma
bandana.elec@gmail.com

[1] Department of Electrical Engineering, Delhi Technological University, Delhi, India

[2] Department of Electronics and Communication Engineering, Bhagat Phool Singh Mahila Vishwavidyalaya, Khanpur Kalan, Sonipat, India

Published online: 31 July 2023

varying weather conditions, minimizing the annual system cost and the loss of power supply probability (LPSP) [15–17]. Navaeefard et al., Bashir et al., and Pirhaghshenasvali et al. used particle swarm optimization (PSO) to determine the optimal size of an HRES [18–20]. Askarzadeh proposed a discrete harmony search (DHS) based algorithm for sizing the SPV and WT in an HRES for a remote Montana State. The proposed technique is simple, handles discrete problems effectively, and obtains solutions rapidly. The suggested approach yielded encouraging findings in a fraction of a second [21]. Ismail et al. presented a hybrid system that combined solar panels and batteries to meet the energy needs of a small rural community in Palestine. The results demonstrate that the proposed solution is more advantageous than a utility grid [22]. Maleki et al. employed a harmony search (HS) based strategy to optimize the design of an SPV/WT-based HRES for rural regions in Rafsanjan, Iran. Furthermore, the HRES was compared to a diesel generator regarding overall cost and emissions [23]. To reduce the total NPC, Heydari et al. created an SPV system using biomass as a backup source. They discovered that the established HRES for delivering energy to agricultural wells in Bardsir, Iran was more cost-effective than solar panels or biomass alone [24]. Tabak et al. researched to optimize a hybrid system comprising an SPV, WT, and biomass generator (BMG) and scaled it to meet the energy needs of a faculty building using GA. Unlike earlier research, the gray wolf optimization (GWO) technique was used for the size optimization of HRES based on reliability. A comparison is then made with the GA results and found more appropriate [12]. In a study by Abu-Hamdeh et al., multivariable linear regression was used to evaluate the effects of different HRES configurations on the COE and the reliability. The observed results are also contrasted with PSO findings and found more appropriate [25]. Diab et al. proposed a simulation model for an HRES microgrid with battery. The optimal size of the proposed system reduces the COE while boosting the reliability and efficiency of the HRES. The optimized HRES was designed using various novel optimization techniques (OT) such as Whale Optimization Algorithm (WOA), Water Cycle Algorithm (WCA), and Moth-Flame Optimizer (MFO). The WOA outperformed the other algorithms in the simulation [26]. Liu et al. proposed an HRES with a high-capacity offshore wind farm and pumped hydro storage (PHS). This work optimized the PHS system size using dynamic economy-based multi-objective technical and economic evaluation [27]. Ganesan et al. suggest a hybrid strategy for an intelligent grid energy management system to minimize fuel expenses, operating and maintenance expenses, hourly power variation, and required load demand using a combination of artificial cell swarm optimization and vapor–liquid equilibrium [28]. Using the PSO approach, Anand et al. determined the optimum size of an HRES for the electrification of rural areas. The grid-connected HRES was found to be the most influential architecture in this investigation. An hourly simulation was used to determine the optimal size of the HRES [29]. The same authors researched to select the ideal size of HRES components using an HS technique. The feasibility analysis of a grid-connected HRES was investigated in this study. The system under discussion has a total NPC that is kept to a minimum by setting boundary limitations, limiting battery storage capacity, and using available land [30]. Saebi et al. researched to find the optimal HRES design for government institutions under attractive policies. The findings suggest that investing in the proposed design will benefit the University after six years [31]. Rajesh et al. proposed a control scheme to improve RES power quality by combining Improved Bat and MFO algorithms [32]. Bakhshaei et al. surveyed to determine the best grid-purchased electricity in grid-connected HRES. The system's running costs were minimized while maintaining a specified degree of stability and operating limitations. Simulation findings demonstrate that determining the best power exchange and incentive rate lowers operational costs [33]. Jamshidi et al. researched to substitute wind speed (W_S) and GHI data with averaged site meteorological data such as atmospheric temperature, elevation, and moisture content. One hundred-five Iranian locations were used as training data to create and test HRES [34]. Rezaei et al. researched grid-independent HRES comprising SPV, WT, and renewable H_2-based Fuel cells for co-supplying electricity and heat to a tourist village in Iran's Isfahan province [35]. Acharya et al. used Refraction-based WOA to solve combined economic emission dispatch challenges by determining the optimum HRES production output. The WT might well be linked to the thermal power plant in the HRES because it cannot be utilized alone to reduce economic and pollution expenses. The proposed Refraction-based WOA strategy was compared with other established approaches regarding financial costs and emissions [36] and revealed appropriate. Das et al. researched the size of an HRES system, which includes solar panels, WT, batteries, and converters, to meet the needs of a rural town. The power generated was used to fulfill the thermal requirements of the community. The HRES optimized for electric and thermal loads compared to identical systems optimized only for electric loads. A comparative study examined the impact of various technical and economic characteristics of the HRES components. The GA optimization approach was also applied to optimize the HRES system [11]. Mishra et al. applied DHS to develop an optimal grid-tied HRES containing SPV and BMG along with a battery grid-linked HRES for supplying power to the selected rural areas [37].

Uncertainty regarding RES presents itself as one of the particular challenges faced in the design of HRES, resulting in various approaches to tackle this issue. Lei et al. provided a bibliographical assessment of current research and advancements in W_S and power forecasting [38]. A unique hybrid

technique is suggested and compared by Catalao et al. for short-term wind forecasting employing wavelet transform [39]. The optimization approaches mentioned in the literature are often based on the previous year's W_S and GHI data. However, it is well recognized that meteorological data from prior years vary dramatically from current estimates. Several models, including uncertainty, have been developed to address this issue to improve energy efficiency. In this context, Khalid, Savkin, and Tascikarao glu et al. employed wind power prediction to optimize HRES, whereas Hocaoglu et al. investigated the influence of GHI forecasting on optimal system design efficiency [40–42]. Zhang et al. put forth a novel framework to determine the optimal size of a grid-connected HRES that comprises a new algorithm as a mix of chaotic search, harmonic search, and simulated annealing. Weather and load forecasts were included using an artificial neural network to increase the precision of the size-optimized design [43]. Gupta et al. suggested the biogeography-based optimization algorithm for size optimization and artificial neural network for solar-wind forecasting. The authors examined the benefits of HRES size optimization using forecasted data rather than historical data [44]. Maleki et al. presented an HS and a hybrid of HS and chaotic search. This study evaluated an HRES that uses iterative neural networks to anticipate the weather data and meet household water needs [45]. Alfonso et al. provide a sizing methodology that utilizes long short-term memory cells to anticipate long-term weather conditions [46].

A thorough review of the literature demonstrates that it is crucial to evaluate various technical, economic, and external energy sources when combining them to achieve the best configuration of an HRES. Intelligent strategies have recently gained popularity and delivered outstanding outcomes. Simulation tools such as the Hybrid optimization model for electric renewable have been used to conduct most studies about grid-connected situations. In contrast, more research on grid-integrated HRES that use intelligent methods must be done. Although biogas and biomass is a significant and potentially valuable resource, particularly in rural regions, power generation based on these resources have received little attention from researchers. It has also been noted that most researchers have used data from just one day to illustrate how well intelligent approaches can solve optimization problems without statistical analysis. Owing to this situation, there is a research gap and a need for more specialized work in HRES. An optimal HRES should be selected based on the RES available at a specific location, considering several technical and economic factors to address this gap. Further, the size optimization approaches presented in the literature are based on the W_S and GHI data collected for previous years. Despite this, it is well-established that the meteorological data from earlier years contain considerable deviations when contrasted with the present values. To the best of our knowledge, none of the studies that have been published in the past have used the ML approaches, viz. GPR, SVR, XGBT, and DT to forecast solar, wind, and temperature data to size HRES.

To address above mentioned issues, in the present work, firstly, hourly forecasting of GHI, temperature, and W_S for one year is carried out using ML approaches, namely SVR, GPR, XGBT, and DT, and found the most appropriate solution. Further, the forecasted GHI, temperature, and W_S data have been used to develop the optimal HRES. The size optimization has been performed using newly developed optimization approaches, viz. CPA, TSA, and AO algorithms. The obtained results are compared and found the optimal solution based on the minimization of NPC. Finally, the result obtained with and without forecasted data has been compared, and found the appropriate solution.

The novelty of this research lies in the application of novel approaches: TSA, CPA, and AO for sizing of HRES along with advanced machine learning (ML) techniques to address the dynamic nature of RES by incorporating precise weather predictions, thereby significantly reducing COE and making the proposed hybrid system more economical.

The main contributions of the work described in this paper can be summarized as follows:

- Novel CPA, TSA and AO algorithms were applied for the first time for tackling the problem of optimal sizing of HRES components.
- To improve the accuracy of size optimization of HRES components, the sizing of HRES is done using forecasted data which is rare in literature. In view of this, various ML models have been used for forecasting the solar irradiation, temperature and wind speed.
- Although being significant and potentially useful resources, particularly in rural regions, biogas and biomass-based power generation have not received much attention from researchers. Addressing this research gap the HRES developed in this research includes BMG and BGG with scheduling of their operation.

The various steps for finding the optimal size of an HRES with meteorological data and forecasted data used in this work is as follows:

- The first phase is collecting data. National Renewable Energy Laboratory was used to acquire the weather data. The meteorological variables GHI, W_S, and temperature were acquired.
- The next stage describes the dataset. Our dataset for this study includes hourly historical data over the past five years (from 2016 to 2020) for GHI, temperature, and W_S. In addition, an examination of the daily distribution of meteorological parameters is carried out, which justifies the removal of values that are either zero or irrelevant.

- The third phase is to forecast those, as mentioned earlier, three meteorological variables using four techniques: GPR, SVR, XGBT, and DT.
- The fourth stage compares the performance of GPR, SVR, XGBT, and DT to identify the best FM.
- The fifth step is to determine the optimum size. Mathematical modeling for this purpose is explained in Sect. 2. Using TSA, CPA, and AO, the appropriate size of the HRES component is determined using historical meteorological data from 2016 to 2020, followed by forecasted solar, wind, and temperature data using the GPR approach.
- The findings from the forecasted models are compared with the historical meteorological data in the sixth phase of the process.
- Finally, the size optimization of HRES has been performed using recently developed TSA, CPA, and AO.

2 Material and methodology

The proposed grid connected HRES comprises of SPV modules, WT, biomass generator (BMG), biogas generator (BGG), battery system, a bi-directional converter, and electrical load. The architecture of the proposed grid connected HRES is illustrated in Fig. 1. The mathematical modeling of the suggested system requires solar irradiance, wind speed, temperature, and load demand information of the selected site. Because of this, the selected site's resource information and load demand have been presented in the subsequent sections.

2.1 Description of selected site

The area under consideration is situated in the Mewat region of Haryana, India. Geographically, it is at the latitude of 28.150N—28.39 0N and longitude of 76.94 0E—77.04 0E. The geographical representation of selected site is illustrated in Fig. 2.

2.2 Resource information of the selected site

A thorough analysis was conducted to determine the potential of RES for the chosen site. The study area has vast potentials for solar energy, wind energy, biomass, and biogas with annual potentials of 1919.90 kWh/m^2/year, 1086.24 kWh/m2/year, 731,818.18 kWh/year, and 172,280 kWh/year, respectively. The total energy requirements of the residents of the selected area could be satisfied using these resources. Figure 3 illustrates the monthly daily average solar irradiation for the selected site. The solar irradiance varies throughout the year, with May having the most significant value of 6.74 kWh/m^2/day and December having the lowest value of 3.53 kWh/m^2/day. Figure 4 illustrates the monthly hourly average ambient temperature of the selected site. The annual range of the ambient temperature has been recorded as 9 to 46 °C. Figure 5 shows the monthly variations in the hourly wind speed at the chosen site. With a calculated wind power density of 124 W/m^2, the anticipated yearly average wind speed is 5.18 m/s.

2.3 Electrical load demand assessment

The load demand for an HRES is crucial when sizing such a system. Therefore the electrical load demand has been determined for the electrification of the proposed rural site. This was carried out by considering the standard of living of the inhabitants and the various electrical appliances they would use. The temperatures of the study area vary between 4 to 46 0C. The fluctuation in the average temperature affects the energy consumption pattern of various home appliances. For the sake of this investigation, the entire year has been split into three distinct seasons, the summer season (April–July), the moderate season (August–November), and the winter season (December–March) also the load demand of these seasons are computed as 3638.955kWh/day, 2849.355kWh/day, and 1533.149 kWh/day respectively. The total load of the study area is estimated as 904,802.9 kWh/year. Figure 6 depicts the daily energy demand for the research area during the three seasons.

2.4 Modeling of HRES components

Mathematical modeling plays a crucial role in developing an optimal design and sizing of HRES. It serves as an imperative tool for understanding the operating performance of each component within the proposed HRES under various conditions. The mathematical modeling of each element within the HRES is outlined as follows:

2.4.1 SPV

The power generated by the SPV system is given by the following equation:

$$P_{pv}^{o} = d_f \times p_{pv}^{r} \times \left(\frac{s^r}{s_r^r}\right) \times [1 + k_t(C_t - C_t^r)] \quad (1)$$

where the SPV power output is represented by P_{pv}^{o} in watts; s_r^r is the GHI at the reference; derating factor is denoted by d_f; p_{pv}^{r} is the rated power. GHI is represented by s^r. The reference cell temperature is defined by C_t^r; the temperature coefficient is denoted by k_t and the cell temperature is represented by C_t.

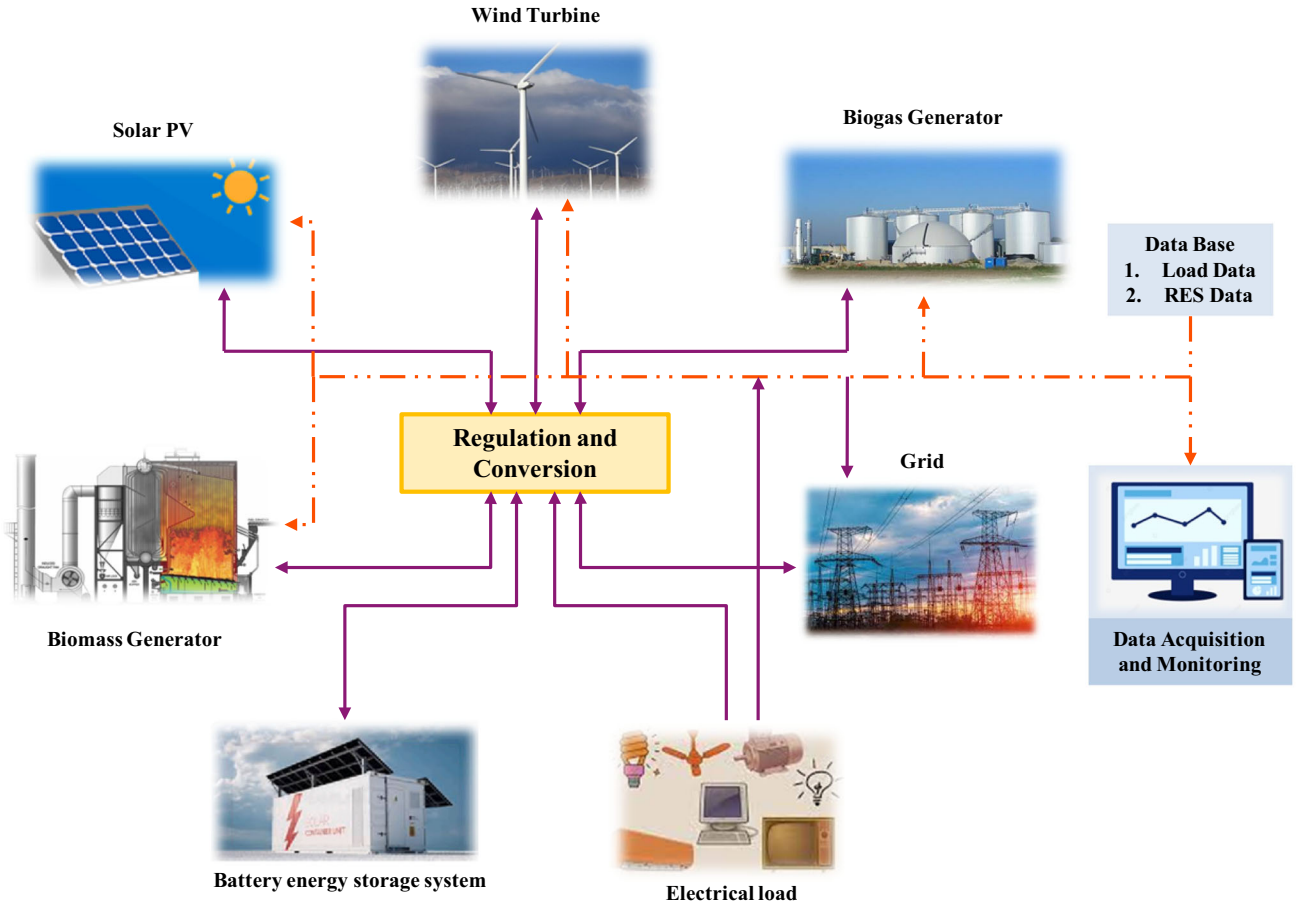


Fig. 1 Architecture of proposed HRES

The energy generated by the SPV system ($PV_E^O(t)$) can be calculated as follows:

$$PV_E^O(t) = P_{pv}^o \times \Delta t \tag{2}$$

2.4.2 WT

The output power of a WT can be calculated using the below mentioned equation [47, 48]:

$$P_{wt} = \frac{1}{2}\rho C_p^{wt} V_w^3 A_r^{wt} \tag{3}$$

where; P_{wt} is the WT power in watts; ρ is the air density, which is equal to 1.225 kg /m^3, the WT power coefficient is denoted by C_p^{wt}; the rotor disk area in m^2 is represented by A_r^{wt} and W_S in m/s at the hub elevation is denoted byV_w^3.

The electricity that is produced by a WT system may be calculated as follows:

$$W_E^o(t) = N_{wt} \times P_{wt} \times \Delta t \tag{4}$$

where; Number of WT is simply represented by N_{wt}.

2.4.3 BMG

The BMG output ($P_{BMG}^O(t)$) can be given by the following equation as [30]:

$$P_{BMG}^O(t) = 3.185 \times 10^{-3} \times (Q_{BM} \times C_{BM} \times \eta_{BM})/(O_{BMG}^h) \tag{5}$$

where; Q_{BM} denotes the available biomass, the biomass calorific value marked by C_{BM}; and the conversion efficiency of BMG is denoted as η_{BM}; operating hours of BMG are denoted by O_{BMG}^h. The energy generated by BMG ($BMG_E^O(t)$) can be calculated as:

$$BMG_E^O(t) = P_{BMG}^O(t) \times \Delta t \tag{6}$$

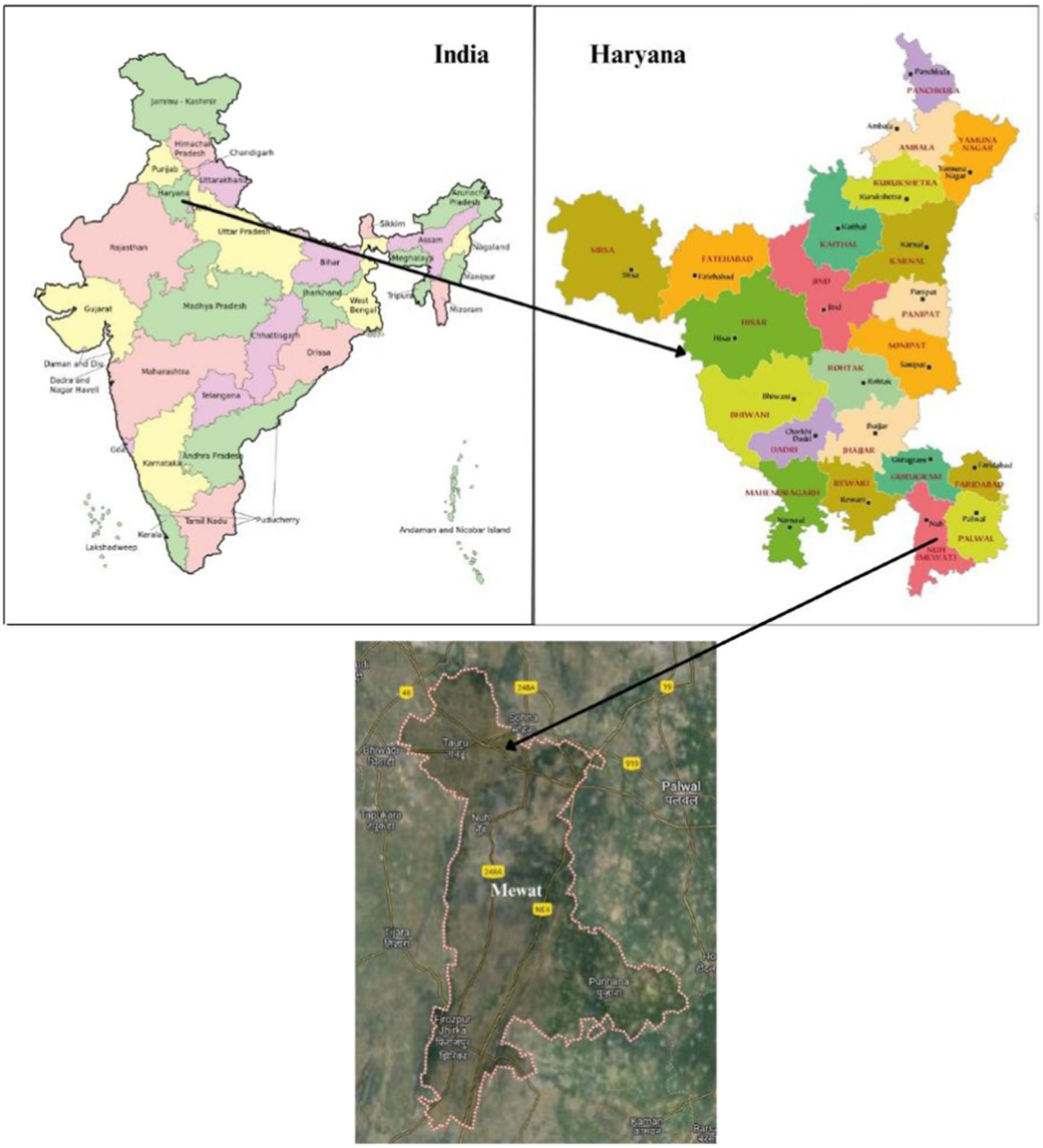


Fig. 2 Geographical representation of the study area

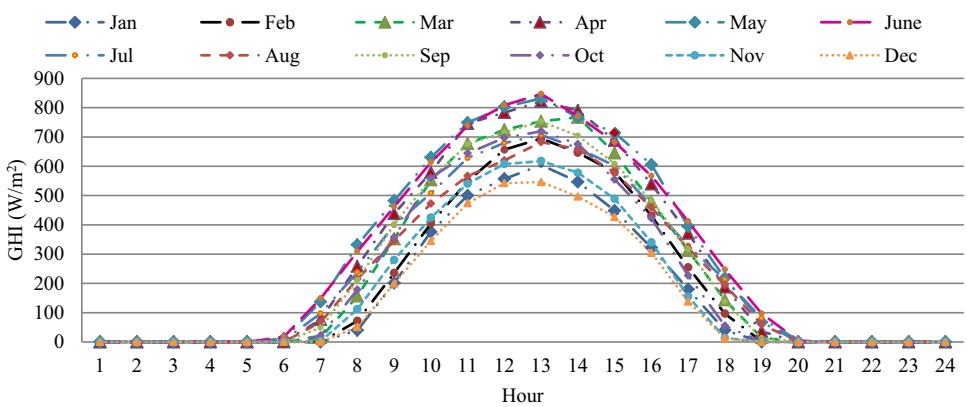


Fig. 3 Monthly hourly average GHI of the selected site

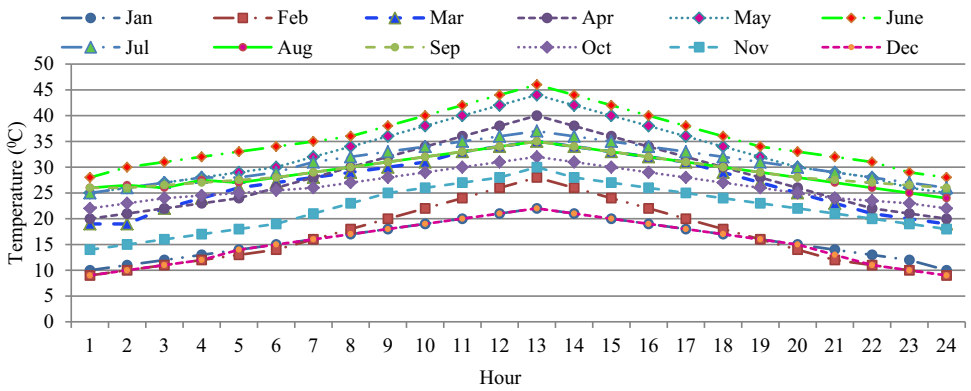


Fig. 4 Monthly hourly average temperature of the selected site

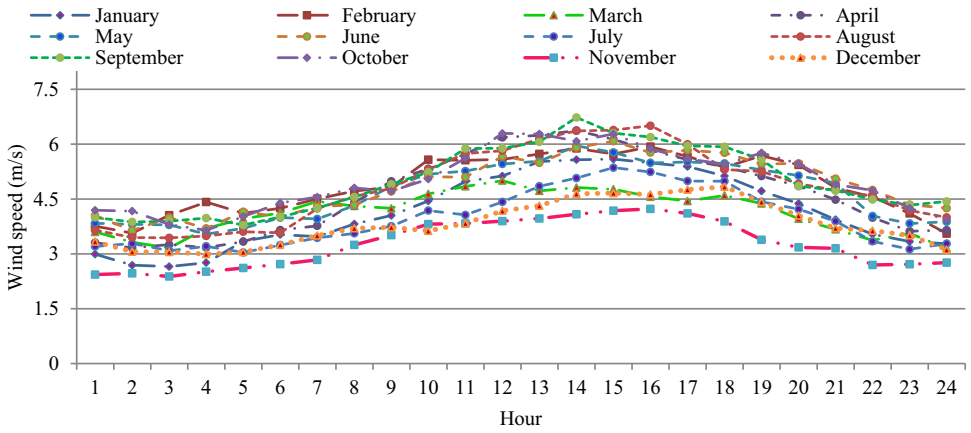


Fig.5 Monthly average wind speed on hourly basis

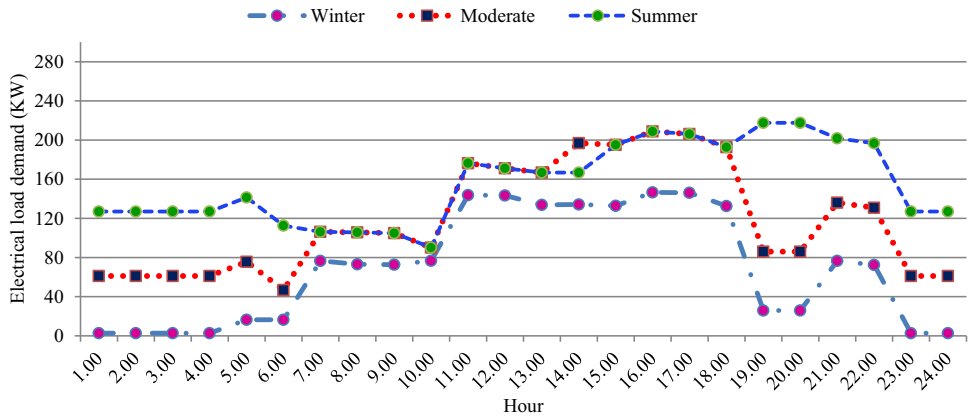


Fig.6 Load demand of selected site

2.4.4 BGG

The output power of BGG ($P^{O}_{BGG}(t)$) can be calculated as [1]:

$$P^{O}_{BGG}(t) = (Q_{BG} \times C_{BG} \times \eta_{BG})/(O^{h}_{BGG}) \quad (7)$$

The availability of biogas/day is indicated by Q_{BG}; biogas calorific value is marked byC_{BG}; BGG conversion efficiency is represented byη_{BG}; and BGG/day operating hours are denoted byO^{h}_{BGG}. The generated energy from BGG system ($BMG^{O}_{E}(t)$) can be computed as:

$$BGG^{O}_{E}(t) = P^{O}_{BGG}(t) \times \Delta t \quad (8)$$

2.4.5 Battery

The surplus energy produced by the RES is stored in a battery which can be utilized as a backup source when needed. The capacity of battery can be calculated based on the energy demand. Surplus power is stored within the battery when the generated energy exceeds the energy demand. The battery

capacity during the charging $E_B^C(t)$ process can be calculated using the following equation [30]:

$$E_B^C(t) = \left(1 - \gamma_B^{SDr}\right) \times E_B^S(t-1) + [E_X^W(t) + E_X^{PV}(t) + E_X^{BMG}(t) + E_X^{BGG}(t)] \times \eta_B^C \quad (9)$$

The energy stored in the battery is marked by $E_B^S(t)$; discharging rate is represented by γ_B^{SDr}; Extra generated energy by SPV, BMG, WT, and BGG systems are defined by $E_X^{PV}(t)$, $E_X^{BMG}(t)$, $E_X^W(t)$, $E_X^{BGG}(t)$ respectively, and the charging efficiency is denoted by η_B^C. When energy demand is greater than generated energy, the energy shortfall is compensated by battery. The battery capacity during discharge ($E_B^D(t)$) can be computed as [30]:

$$E_B^D(t) = \left(1 - \gamma_B^{SDr}\right) \times E_B^S(t-1) - \frac{E_{df}(t)}{\eta_B^c \times \eta_B^d} \quad (10)$$

where; η_B^d indicates the discharging efficiency; charging efficiency is denoted by η_B^c; and $E_{df}(t)$ represents an unmet energy demand.

2.4.6 Grid

When energy demand > generated energy and the minimum battery capacity ($E_B^{mn}(t)$), energy will be drawn from the grid and can be determined as follows [18].

$$E_g^p(t) = E_D(t) - \left[W_E^O(t) + BMG_E^O(t) + BGG_E^O(t) + \left[PV_E^O(t) + \left(E_B^S(t) - E_B^{mn}(t)\right) \times \eta_{inv}\right]\right] \quad (11)$$

where $E_g^p(t)$ is the energy drawn from the grid; energy demand is denoted by$E_D(t)$; Inverter efficiency is indicated by η_{inv}.

Equation (12) can be used to determine the energy sold to grid ($E_g^s(t)$)

$$E_g^s(t) = \left[W_E^O(t) + BMG_E^O(t) + BGG_E^O(t) + \left[PV_E^O(t) - \left(E_B^{mx}(t) - E_B^S(t)\right) \times \eta_{inv}\right]\right] E_D(t) \quad (12)$$

Maximum capacity of battery is marked by $E_B^{mx}(t)$.

2.5 Problem definition

The HRES optimal size optimization problem is defined and discussed in this section. We can determine the component capacities that result in the most cost-effective and reliable HRES by specifying the size optimization problem. The optimization process considers an objective function, decision variables and a number of relevant constraints described in the following paragraphs.

2.5.1 Objective function

For size optimization of HRES, the main objective function is NPC, which includes the NPC of all the components involved in the HRES design [1].

The total NPC of HRES (N_{PC}^T) may be expressed as follows:

$$N_{PC}^T = NPC_{PV} + NPC_{BMG} + NPC_{inverter} + NPC_{Battery} + NPC_{BGG} + NPC_{WT} - \in_S^G + \in_P^G \quad (13)$$

NPC of SPV system is denoted by NPC_{PV}; NPC_{BMG} is the NPC of BMG system; $NPC_{inverter}$ represents NPC of inverter; $NPC_{Battery}$ defines NPC of battery; NPC of BGG is indicated byNPC_{BGG}; NPC_{WT} is the NPC of WT ; $\in_S^G$ is the cost of energy sold to grid and cost of energy purchased from grid is denoted by$\in_P^G$.

2.5.2 Constraints

While minimizing total NPC, the optimized HRES system should also satisfy the boundary constraints. Further, the sizes of the system components may vary in the framework designed to satisfy load requirements. Therefore, the boundaries for various HRES components are defined as follows [1]:

$$N_{pv}^{min} \le N_{pv} \le N_{pv}^{max} \quad (14)$$

$$N_{Battery}^{min} \le N_{Battery} \le N_{Battery}^{max} \quad (15)$$

$$P_{BMG}^{O\ min} \le P_{BMG}^O \le P_{BMG}^{O\ max} \quad (16)$$

$$P_{BGG}^{O\ min} \le P_{BGG}^O \le P_{BGG}^{O\ max} \quad (17)$$

$$P_{wt}^{min} \le P_{wt} \le P_{wt}^{max} \quad (18)$$

$$E_B^S(t)^{min} \le E_B^S(t) \le E_B^S(t)^{max} \quad (19)$$

where; N_{pv} indicates the number of PV panels; the number of batteries is represents by $N_{Battery}$.

Reliability The LPSP is a constraint-based on a reliability indicator. It is an indicator that measures the proportion of energy that cannot be provided to the load, as shown in Eq. 20. It accepts a value between 0 and 1 or 0% and 100%. A 0% LPSP indicates that the energy demand is entirely satisfied and reliability is at its highest level. The system could not energize the load and was unreliable when LPSP was

100%.

$$\mathrm{LPSP} = \sum_{t=1}^{t=8760} \left(\frac{Unmet\ load\ for\ one\ yea}{Total\ load\ for\ one\ year} \right) \tag{20}$$

Renewable fraction (RF) The RF is a measure of the electricity generated by RES in a system and supplied to the load, is described in the following equation [1]:

$$\mathrm{RF} = 1 - \frac{\text{Energy generated by RES}}{\text{Energy supplied to load}} \tag{21}$$

2.6 Optimization techniques

The recently developed optimization techniques CPA, TSA, and AO which are used for size optimization of HRES have been discussed in the subsequent sections.

2.6.1 Colony predation algorithm (CPA)

The CPA is a novel stochastic optimization approach developed by Chen et al., the behavior of social animals serves as the basis for this algorithm. The main focus herein is on colony predation, a strategy used by animals to evade predators and boost hunting success rates. Wolf packs, hyena packs, lion packs, piranha packs, and other animals that hunt in packs frequently use this tactic. Predators frequently obtain additional prey through colony predation, improving individual survival rates. This is a strategy used by animals that live in colonies to ensure their survival and maximize the effectiveness of their hunt for food. Hyenas and wolves, for instance, communicate and work together through colony predation to capture more prey than they could individually. The two most popular strategies for increasing the likelihood of successful hunting are dividing and encircling animals. Another tactic animals use in circumstances when consumption surpasses their produce. In this behavior, they will switch to another prey, increasing their predation's effectiveness. CPA draws inspiration from collective hunting; a strategy herd animals use to increase individual survival rates. CPA also takes into account the natural phenomena of survival of the fittest [49].

Step1: Communication and collaboration

Predation success rates are higher for animals that hunt in groups, because they communicate and work together more efficiently.

Using the following formulae, we can describe how individuals communicate and hunt for food in groups [49].

$$\mathrm{Y_n^m}(\tau+1) = \mathrm{Y_n^{m_1}}(\tau) + (1-p).\left(\frac{\mathrm{Y_1}(\tau) + \mathrm{Y_2}(\tau)}{2} \right) \tag{22}$$

where; p is the random number lies between 0 to 1; the individual looking for food is marked by $\mathrm{Y_n^{m_1}}(\tau)$; the two closest positions to prey in the $\mathrm{n^{th}}$ dimension are indicated by $\mathrm{Y_1}(\tau)$and$\mathrm{Y_2}(\tau)$; $n\epsilon$ 1,2,…., dim; dim is the dimension of population; the individual's latest updated position is represented by $\mathrm{Y_n^m}(\tau+1)$.

Step2: Disperse food

The initial approach in colony prediction separates the prey from its group by driving it in various directions. The predation strategies of the individuals in the search were mathematically modeled and as follows:

$$\mathrm{Y}(\tau+1) = \mathrm{Y_b} - \mathrm{P_s}.(\mathrm{k_1}.(u_b - l_b) + l_b) \tag{23}$$

where the position of a population is denoted by Y (τ + 1), the position of food is marked by $\mathrm{Y_b}$; $\mathrm{P_s}$ denotes the strength of prey, $\mathrm{k_1}$ represents [$\mathrm{K_1}$; $\mathrm{K_2}$; $\mathrm{K_3}$; …$\mathrm{K_n}$], which is a random number between 0–1;u_b, l_b are the upper and lower bounds respectively.

Step3: Encircle food

The second tactic is used by the hunting party to encircle and approach animals. This step can be mathematically mimic using the following equation:

$$\mathrm{Y}(\tau+1) = \mathrm{Y_b} - 2 \cdot \mathrm{P_s} \cdot \mathrm{d} \cdot \mathrm{e}^{l} \cdot \tan\left(\frac{\pi}{4} \cdot l \right) \tag{24}$$

where the distance that separates the current member from the prey is represented by d; the random number is defined by l which lies between 0 and 1. The value of d is calculated as follows:

$$\mathrm{d} = |\mathrm{Y_b} - \mathrm{Y}(\tau)| \tag{25}$$

The current hunter population is marked by Y(τ).

Step4: Supporting closest individual

Recognizing the group's potential challenges in pursuing prey, the closest individual requests peer assistance, which can be described as follows.

$$\mathrm{Y}(\tau+1) = \mathrm{P_{nearest}} \tag{26}$$

where, location of the nearest predator is marked by $\mathrm{P_{nearest}}$.

Step 5: Searching for the food

If no prey is located nearby or if the food supply is too far away from the prey, the surviving individuals will seek another food source. In the following way this behavior may be expressed as follows:

$$\mathrm{d_1} = |2 \bullet \mathrm{k_4} \bullet \mathrm{Y_{rand}} - \mathrm{Y}(\tau)| \tag{27}$$

$$\mathrm{Y}(\tau+1) = \mathrm{Y_{rand}} - \mathrm{P_s} \bullet \mathrm{d_1} \tag{28}$$

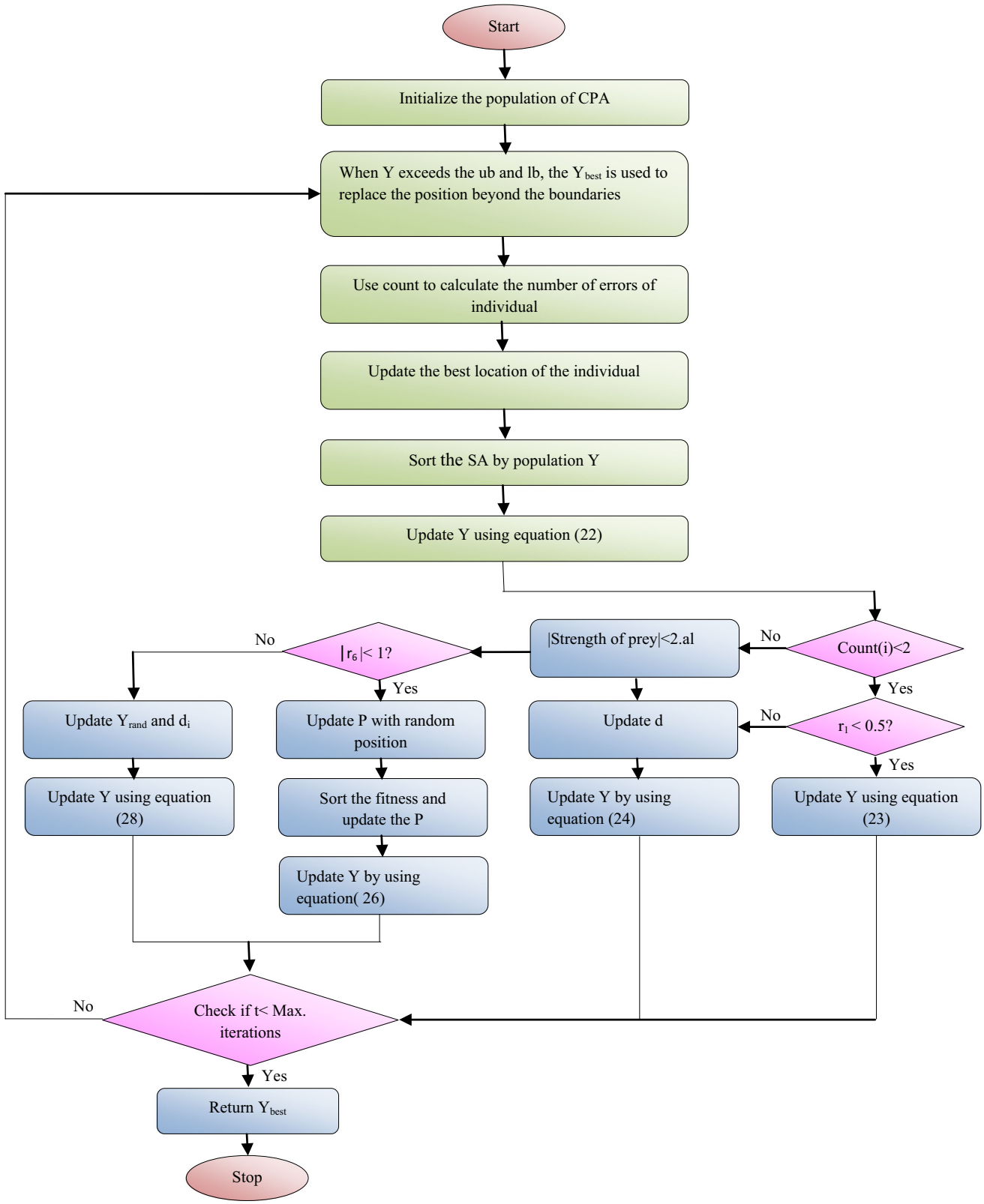


Fig. 7 Flowchart of CPA algorithm

where; random number is defined as k_4 and its value lies between [0 and 1]. An individual's random position is indicated by Y_{rand}; d_1 represent the distance of random group movement. Further the flowchart for proposed CPA is presented in Fig. 7.

2.6.2 Aquila optimization (AO) algorithm

In 2021, Abualigah et al. introduced the population-based meta-heuristic algorithm known as the Aquila Optimizer (AO). The eagle, known in Latin as Aquila, is a bird of prey that hunts by using its quickness and agility to seize its prey. It has strong feet and sharp talons are considered as the most intelligent and proficient hunting tools. Aquila has four main hunting techniques: high soar with a vertical stoop, contour battle with a short glide, low fly with a gradual descent, and strolling and catching prey. The primary method of hunting used by Aquila is narrow exploration. Quick soaring moves were used to assault the target's body after landing in the designated area and circling the target. Aquila's four core hunting strategies served as inspiration for the development of the Aquila Optimizer (AO), a nature-inspired optimization algorithm that fundamentally explains the hunt's various stages. The various hunting steps used in AO algorithms are explained below [50].

Step1: Expanded exploration (Z1): high soar with a vertical stoop

While hunting for food, Aquila utilizes expanded exploration, which involves flying at a high altitude and simultaneously looking over a wide region. When Aquila has located its target, it will do a vertical dive to capture it successfully. As a direct outcome of this, the mathematical representation of this behavior is as follows: [50].

$$Z_1(\varphi+1) = Z_b(\varphi) \times \left(1 - \frac{\varphi}{T}\right) + (Z_m(\varphi) - Z_b(\varphi) \times r_1) \tag{29}$$

where;φ and T are the current and maximum iteration respectively; $Z_1(\varphi+1)$ is the solution of the next iteration of φ, $Z_b(\varphi)$ denotes the best position, $Z_m(\varphi)$ is the average location of all aquilas in the current iteration; n is the population size; and r_1 is random number between 0 and 1.

Step2: Narrowed exploration (Z2): contour flight with short glide attack

While hunting, Aquila uses the most common strategy called "narrowed exploration." In this strategy, upon landing inside the prescribed region and its subsequent flight around the target, it launches a series of quick gliding attacks employing short glides against the victim. The formula for the position update can be represented by the notation shown below:

$$Z_2(\varphi+1) = Z_b(\varphi) \times lf(d) + Z_r(\varphi) + (x - z) \times r_2 \tag{30}$$

where; arbitrary location of Aquila is marked by $Z_r(\varphi)$; the dimension size is denoted by d, the random value is represented by r_2 which lies between 0 and 1; x and z are used to present the spiral shape in the search and the Levy Flight Function is marked by lf(d).

Step3: Expanded exploitation (Z3): low flight with a slow descent attack

In the expanded exploitation strategy, Aquila descended vertically to launch its first assault, after an approximate assessment of its prey position. Aquila uses a given area to approach and attack targets. This behavior can be expressed mathematically as follows:

$$Z_3(\varphi+1) = (Z_b(\varphi) - Z_m(\varphi)) \times \ -r_4 + ((ub - lb) \times r_5 + lb) \times \delta \tag{31}$$

where; exploitation adjustment parameters are marked by α, δ; lower and upper bounds are represented by lb and ub; random numbers are between 0 and 1 and denoted by r_4, r_5.

Step4: Narrowed exploitation (Z4): walking and grabbing prey

When pursuing targets in the direction of their rescue paths, the aquilas use a narrow exploitation approach. They then assaulted the target while flying mid-air [51].

$$Z_4(\varphi+1) = qf \times Z_b(\varphi) - (g_1 \times Z(\varphi) \times r_6) - g_2 \times lf(d) + r_7 \times g_1 \tag{32}$$

where $Z(\varphi)$ denotes the current position; qf denotes the quality function value. While tracking prey, the value of the movement parameter g_1 ranged from -1 to 1; g_2 represent the flight slope when chasing prey. The random numbers r_6 and r_7 were chosen to be between 0 and 1. Further the flowchart for AO algorithm is presented in Fig. 8.

2.6.3 Tunicate swarm algorithm (TSA)

TSA is an algorithm that takes inspiration from tunicate colonies. Tunicate is a group of marine animals. They make dwellings on piers, rocks, or the undersides of boats and ships. Most often, they resemble small blobs of color. The tunicate can emit intense blue-green light, or bioluminescence, which makes it visible from a distance. They have two ends: an open end and a closed end. The open end is utilized for propulsion, such as jet propulsion employing atrial siphons. Tunicates move by using bursts of fluid. The force of this explosion is so intense that it can cause tunicates to travel vertically in the water. The shape of this animal is measured in millimeters.

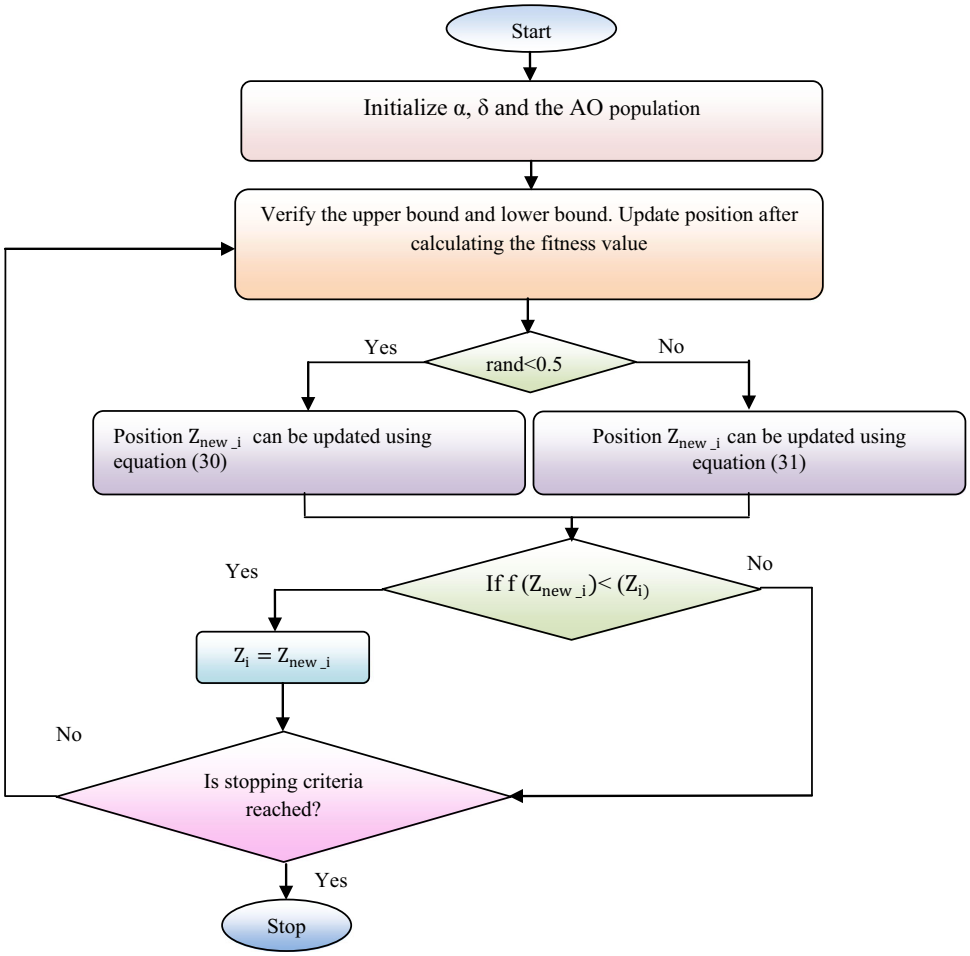


Fig. 8 Flow Chart of AO algorithm

Tunicates have the expertise and ability to locate food sources in the water even when there is no information available on food sources. The ability to identify food is present in tunicates. The most crucial tunicate activities are their jet propulsion and swarm intelligence, which they use to identify the optimal food supply. The mathematical modeling of the initial behavior of the tunicates, which is the jet's propulsion, needs to satisfy three requirements. These requirements aim to avoid conflicts amongst tunicates, modify the potential tunicate's position, and get closer to the potential tunicate. Conversely, the swarm behavior keeps track of fellow seekers' presence to arrive at the most appropriate and optimal solution [51].

Step1: Preventing search agent disputes

To prevent the disputes among search agents (SA), vector $\overrightarrow{B}$ is used to calculate the new search agent location

$$\overrightarrow{B} = \frac{\overrightarrow{H}}{\overrightarrow{N}} \tag{33}$$

$$\overrightarrow{H} = r_2 + r_3 - \overrightarrow{E} \tag{34}$$

$$\overrightarrow{E} = 2.r_1 \tag{35}$$

where, the gravity force is indicated by $\overrightarrow{H}$; $\overrightarrow{E}$ represents deep ocean advection;. r_1, r_2, r_3 denote random values between 0 and 1; $\overrightarrow{N}$ represents the social forces among SA.

The following formula is used to compute the vector N:

$$\overrightarrow{N} = [Q_{min} + r_1(Q_{max} - Q_{min})] \tag{36}$$

where;Q_{min}, Q_{max} represent initial and subordinate speeds to make social interaction.

Step2: Movement towards the direction of best neighbour

After avoiding the disputes between neighbors, the SA moves towards the direction of best neighbor

$$\overrightarrow{P_D} = | \overrightarrow{F_S} - R_{rand} \bullet \overrightarrow{Q_p}(y) | \tag{37}$$

where; the distance between the food source and SA is denoted by $\overrightarrow{P_D}$; the current iteration is represented by y, $\overrightarrow{F_S}$ indicates the position of optimum food source; $\overrightarrow{Q_p}(y)$ is

the position of tunicate; and R_{rand} represents random value between 0 and 1.

Step3: Converge towards the best search agent

The SA can maintain its position towards the optimal location of SA which are near to food source..

$$\overrightarrow{Q_p}(y) = \overrightarrow{F_S} + \overrightarrow{B} \bullet \overrightarrow{P_D}, \text{ if} R_{rand} \geq 0.5 \tag{38}$$

$$\overrightarrow{Q_p}(y) = \overrightarrow{F_S} - \overrightarrow{B} \bullet \overrightarrow{P_D}, \text{ if} R_{rand} < 0.5 \tag{39}$$

where; $\overrightarrow{Q_p}(y)$ represents the updated position of tunicate with respect to the position of food sources, $\overrightarrow{F_S}$

Step4: Swarm behavior

To mathematically imitate the swarm behaviour of tunicates, the first two optimal solutions remain intact, and the locations of the remaining SA are updated based on the optimal location of the SA. To describe the tunicate swarm behaviour, the following formula can be used:

$$Q_P\left(\overrightarrow{y} + 1\right) = \frac{\overrightarrow{Q_P}(y) + \overrightarrow{Q_P}(y+1)}{(2 + r_1)} \tag{40}$$

Further the flowchart for TSA is presented in Fig. 9.

3 Results and discussion

In the current work, the hourly forecasting of GHI, temperature, and W_S is carried out for one year by utilizing ML techniques, specifically SVR, GPR, XGBT, and DT. The obtained results have been compared with the help of statistical indicators, and the best solution has been determined for GPR. Further, the forecasted data obtained through GPR is used for sizing HRES. The sizing of HRES is carried out using the recently developed OT: CPA, TSA and AO using forecasted data and without forecasted data. MATLAB was used for CPA, TSA and AO algorithms to run hourly simulations over one year for a selected grid-linked HRES. The obtained results are compared based on minimal NPC with the reliability of the power supply. In addition, in view of developing HRES, a comparative analysis of various parameters obtained after size optimization using forecasted data and without forecasted data is performed, and the results of TSA with forecasted data are found to be more optimal with a reduction of 0.42% in COE.

3.1 Performance indicators for forecasting

It isn't easy to assess the accuracy of an FM without doing the performance assessment; therefore, statistical indicators have been used for the performance evaluation. The indicators that are employed for comparative research are termed Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Correlation Coefficient (R^2). The MAE measure is useful for determining FM prediction errors. The correlation coefficient reveals how closely two models are related to each other, whereas the RMSE indicator analyzes the model's overall accuracy.

3.2 Training and testing phase

The different FMs are trained using four years of data, i.e., 2016–2019 of wind, GHI and temperature. The training phase of FM offers a better understanding of the model's accuracy and is essential for selecting the best FM. Further, in the testing phase, also known as the real prediction stage, the GHI, temperature, and W_S are forecasted using the model effectively trained in the previous phase. To evaluate the effectiveness of the FM, each model uses 2020 data for prediction. The performance indicators for GHI, temperature, and W_S for the training and testing phases, respectively, have been determined and presented in Tables 1, 2, 3 respectively. From Tables 1, 2 and 3, it is apparent that the GPR method performs better for prediction out of all of these ML techniques. The GPR method performs better because the generated statistical indicators have the lowest possible value, while the R^2 has the highest possible value. Tables 1, 2, 3 shows that the GPR model has the lowest RMSE, MSE, and greatest R^2 for all training and testing phases, including GHI, temperature, and W_S. Therefore, the forecasted GHI, temperature, and W_S data obtained by the GPR approach have been used to develop HRES. Further, the prediction performance of the proposed models, along with meteorological or true data, is shown in Figs. 10a, 11a, and 12a. The comparative analysis of the errors for ML FM is presented in Fig. 10b, Fig. 11b and Fig. 12b.

3.3 Optimization results of proposed HRES

After obtaining the forecasted data for GHI, temperature and W_S, in the present work, the optimal grid-linked HRES, involving various RES, was developed to fulfil the energy needs of the given site based on minimal NPC. Three recently developed optimization approaches viz. TSA, CPA, and AO were applied to optimize the size of the selected HRES. Further, several economic input indicators used in obtaining the optimization outcomes are illustrated in Table 4. The hourly simulation for selected HRES has been conducted in MATLAB for one year. To facilitate this analysis, 150 iterations were performed. An Intel® Core TM i5-8250 was used to create and test the recommended optimizers. Further, the yearly real interest, inflation and escalation rates are fixed at 0.11, 0.075, and 0.05, respectively. The parameters of the TSA, CPA and AO algorithms are set as follows: TSA: maximum iteration = 150, SA = 30, CPA: maximum iteration = 150;

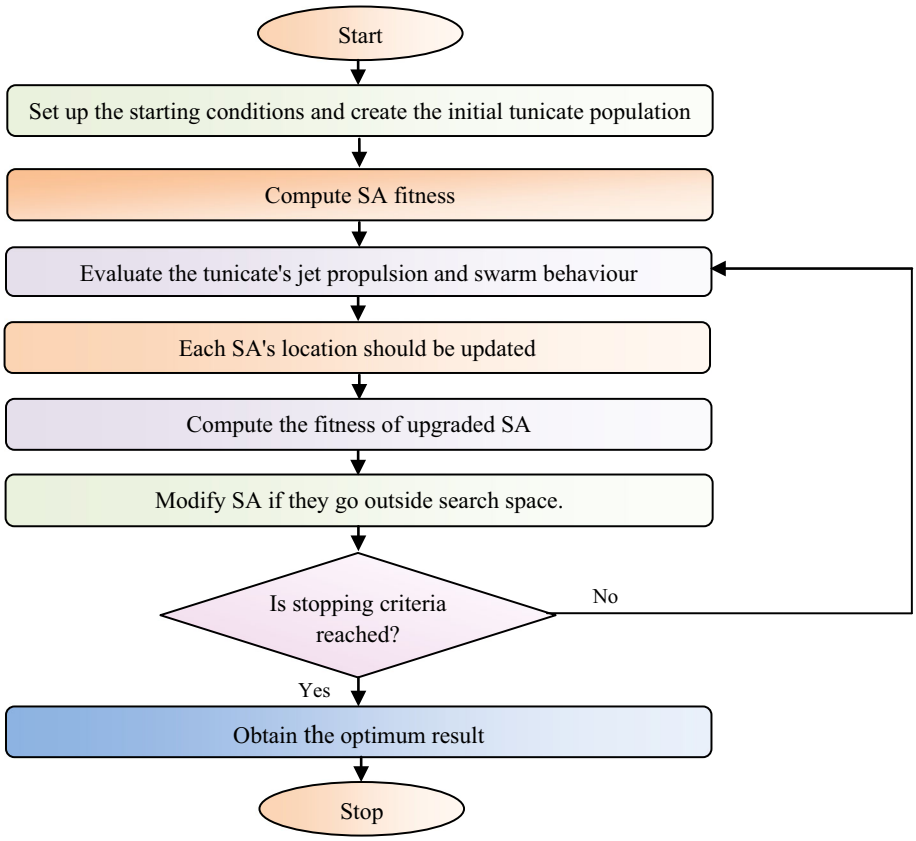


Fig. 9 Flowchart of TSA algorithm

Table 1 Training and testing performance comparison for GHI

ML model	Training phase			Testing phase		
	RMSE	MAE	R^2	RMSE	MAE	R^2
DT	26.28087	20.10835	0.989453	30.149	23.3189	0.986104
SVR	33.83328	26.35761	0.982519	33.469	26.0625	0.982875
GPR	23.70055	18.45589	0.991422	26.613	20.8928	0.989172
XGBT	38.88621	30.73992	0.976908	43.199	34.0433	0.971471

Table 2 Training and testing performance comparison for Temperature

ML method	Training phase			Testing phase		
	RMSE	MAE	R^2	RMSE	MAE	R^2
DT	1.840803	1.358063	0.959016	2.261181	1.735343	0.939188
SVR	2.399271	1.794306	0.930376	2.566462	1.961314	0.921659
GPR	1.775453	1.331063	0.961874	1.801434	1.404804	0.961403
XGBT	2.705954	2.16118	0.91144	2.694554	2.22985	0.913644

dimensions = 5; SA = 30; AO: exploitation adjustment coefficients; $\alpha = 0.1$; $\delta = 0.1$; Population size = 30, maximum iteration = 150.

Various indices obtained by applying TSA, CPA and AO algorithms with forecasted and without forecasted data have been presented in Table 5. By analyzing and comparing the obtained result presented in Table 5, it has been evident that, out of three OTs, TSA with forecasted data gives the most optimal results. It is also found that there is a 0.42% reduction in COE for the TSA model with forecasted data. The optimization result acquired for the selected grid linked HRES with their size using the forecasted data is shown in Table 6.

The contribution of various RES in yearly power production via TSA, CPA and AO has been obtained and illustrated

Table 3 Training and testing performance comparison for W_S

ML method	Training phase			Testing phase		
	RMSE	MAE	R^2	RMSE	MAE	R^2
DT	0.7351721	0.513848662	0.845943	1.1948515	0.88169812	0.836684
SVR	0.8769833	0.626250966	0.871789	1.1217303	0.81555375	0.882985
GPR	0.6439014	0.452467609	0.891567	1.0579799	0.76115085	0.889172
XGBT	0.8308530	0.605577454	0.876908	1.0551265	0.74767111	0.821514

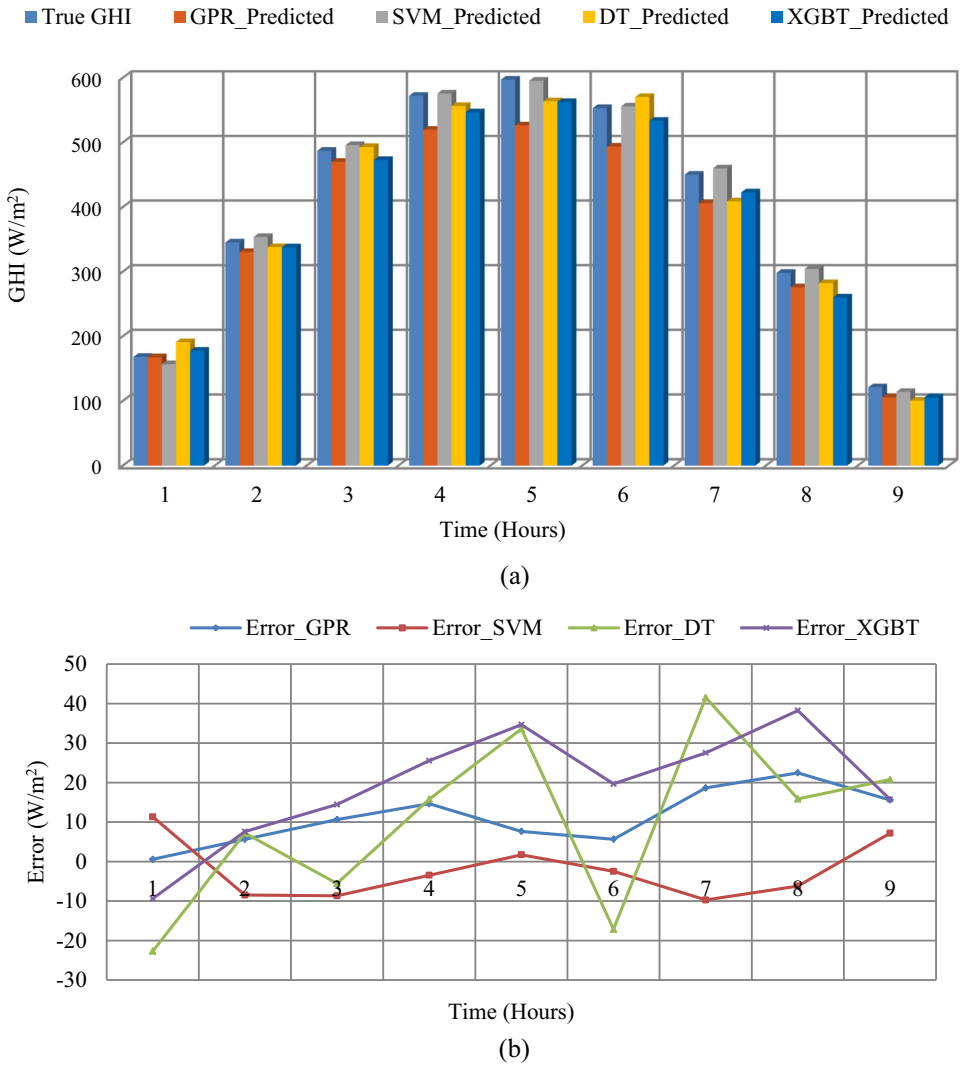


Fig. 10 Hourly forecast performance comparisons. **a** GHI. **b** Error

in Fig. 13. The result demonstrates that the SPV array generated the highest power in all three algorithms.

Furthermore, the contribution of various system elements to the NPC is demonstrated in Fig. 14. It was found that, in the case of TSA, the cost of the WT system was the highest, followed by BGG, SPV, battery, and BMG. However, for the CPA and AO algorithms, the highest share was for BGG, followed by WT, SPV, BMG, and battery. In addition, Fig. 15 shows the energy generated by different HRES components during one day.

A comprehensive statistical analysis for all three proposed techniques is provided in the present work. The best and worst values of the cost function, mean, median, standard deviation (SD), relative error (RE), RMSE, MAE, and efficiency were obtained and presented in Table 7 for CPA, TSA, and AO. As it is evident from Table 7 that the statistical indicators, RMSE,

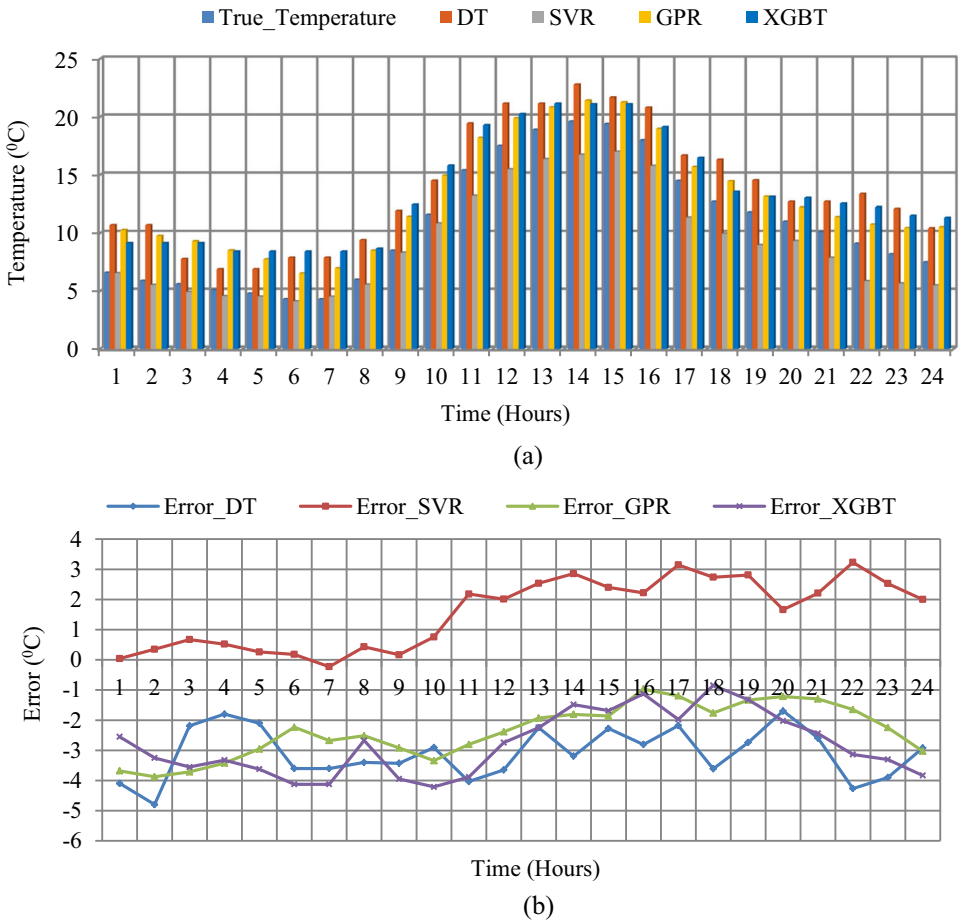


Fig. 11 Hourly forecast performance comparison. **a** Temperature. **b** Error

standard deviation (SD), MAE, and average RE are all minimum for TSA, while efficiency is highest when contrasted with CPA and AO, demonstrating the TSA's superiority over other algorithms.

Figure 16 depicts the convergence curves for the TSA, CPA, and AO algorithms of the proposed grid-connected HRES model. It can be seen that TSA and AO converged and produced a fixed value at iterations 16 and 32, respectively. However, at iteration 60, CPA stabilized at a constant value. Consequently, TSA has a higher convergence rate than AO and CPA.

Furthermore, the findings of the present research are summarized as follows:

- The four recommended ML techniques: GPR, SVR, XGBT, and DT, were used to forecast temperature, W_S and GHI.
- Based on the statistical indicators, the FM was contrasted, and GPR FM was identified as the best FM.
- The data obtained from GPR for GHI, W_S and temperature were used as input data for the size optimization of HRES components.
- The recently developed optimization algorithms: TSA, CPA, and AO were initially used to find the optimal size of HRES components using historical meteorological data. The optimization results were also obtained by TSA, CPA, and AO techniques using the forecasted GHI, W_S and temperature data obtained through the GPR approach.
- The findings of size optimization obtained through the forecasted data were compared with the results obtained for historical meteorological data. The size optimization results obtained using the forecasted data are more appealing due to low NPC and COE.
- Finally, the optimization results of TSA, CPA, and AO algorithms obtained using forecasted data are compared, and it was found that the TSA with forecasted data delivers an optimal solution with a reduction of 0.42% in COE.
- The sizes obtained for the different components of the optimal system include 282 kW of PV, 115.5 kW WT, 10 kW BGG, 25 kW BMG, and 36 kWh BESS.
- For the optimal design, the NPC and COE are estimated to be \$647,800 and \$0.0701/kWh, respectively.
- The share of SPV system is highest in energy generation with 61%, followed by WT, BGG and BMG with 24%, 10% and 5%, respectively.

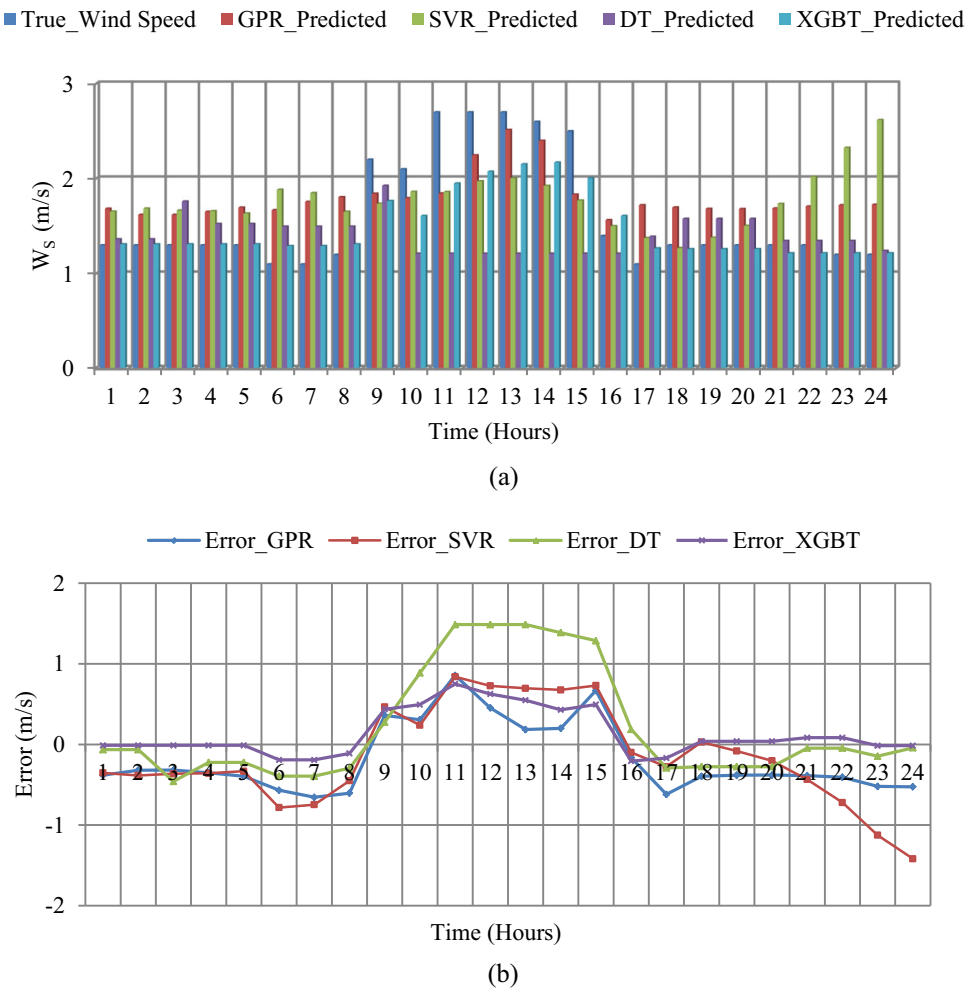


Fig. 12 Hourly forecast performance comparison. **a** W_S. **b** Error

Table 4 Economic parameters of HRES components [30]

Components	Capital cost (CC) ($)	Operational & maintenance cost (OMC) ($)	Salvage cost (SC) ($)	Fuel Cost (FC) ($/tonne)
SPV (235 W)	166.4	3.328	16.64	–
WT (3300 W)	1,077	32.31	323.1	–
BGG (1000 W)	572	28.6	171.6	6.93
BMG (1000 W)	895.267	44.763	268.58	13

4 Conclusion

In this research, the size optimization of HRES has been carried out for the outlying region of Mewat, India. Grid-connected configurations were designed and presented using the forecasted data for GHI, temperature, and W_S. The four ML methods: SVR, DT, GPR, and XGBT, were used for forecasting, out of which GPR performs better. Therefore, data obtained from GPR FM was used for the sizing of HRES. Further, the HRES was optimized using recently developed CPA, TSA, and AO algorithms. All three algorithms performed well and were suitable for the size optimization of the HRES; however, the sizes of the HRES components for the grid-connected scenario have been found most optimal via the TSA algorithm. The optimal HRES obtained from TSA included a 282 kW SPV, a 115.5 kW WT, a 10 kW BGG, a 25 kW BMG, and a 36 kWh battery bank. The COE and overall NPC costs were anticipated to be 0.0701 $/kWh and $647,800, respectively. The findings of this study can be utilized to create HRES for locations with similar geographic characteristics, which can improve the lifestyle of people in

Table 5 Results of Optimization for forecasted data and without forecasted data

Parameters	TSA		CPA		AO	
	With Forecasted data	Without Forecasted data	With Forecasted data	Without Forecasted data	With Forecasted data	Without Forecasted data
Energy generated by SPV (W)	4.670×10^5	4.438×10^5	4.4833×10^5	4.354×10^5	4.46×10^5	4.321×10^5
Energy generated by WT (W)	1.80×10^5	1.651×10^5	1.293×10^5	1.182×10^5	1.40×10^5	1.20×10^5
NPC ($)	647,800	659,923	658,120	664,344	660,831	669,765
COE ($/kWh)	0.0701	0.0705	0.0709	0.0719	0.0728	0.0737
Grid purchase (GP) energy	4.324×10^4	6.798×10^4	5.120×10^4	6.828×10^4	5.420×10^4	6.992×10^4
Grid sale (GS) energy	5.592×10^3	4.435×10^3	5.087×10^3	4.897×10^3	5.016×10^3	4.986×10^3

Table 6 Optimum configuration of grid linked HRES using TSA, CPA and AO algorithm

OT	N_{pv}	N_{wt}	$N_{Battery}$	PV_E^O (kW)	P_{wt}(kW)	E_B^S (kWh)	P_{BMG}^O (kW)	P_{BGG}^O (kW)	NPC ($)	COE ($/kWh)
TSA	1200	35	15	282	115.5	36	25	10	647,800	0.0701
CPA	1000	28	14	235	92.4	33.6	30	12	658,120	0.0709
AO	1010	25	18	237.4	82.5	43.2	30	15	660,831	0.0737

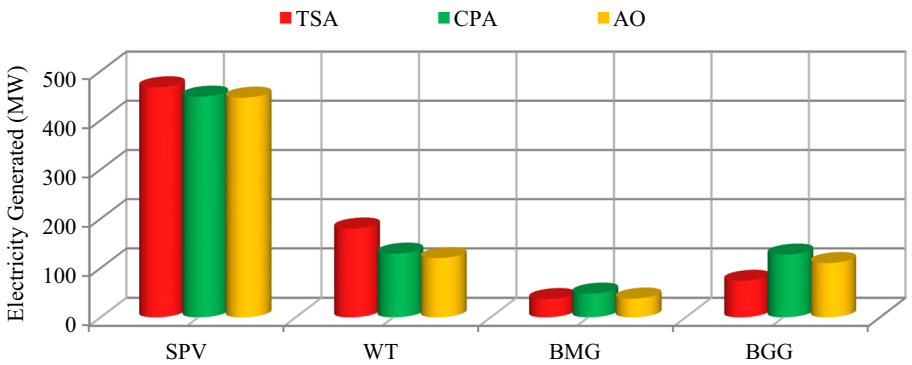


Fig. 13 Total yearly energy generation for various OT

rural areas by providing an uninterrupted power supply. This research can also help to some extent to fulfill the goal of the Indian government's "power to all" with its availability 24 × 7.

The current research could be extended in the future by including the following:

- Other regions in various towns, districts, states, and other administrative divisions with areas that contain hydro resources in addition to solar, wind, and biomass can be researched.
- The proposed FM also benefits future possibilities for smart grid demand response applications and energy management.

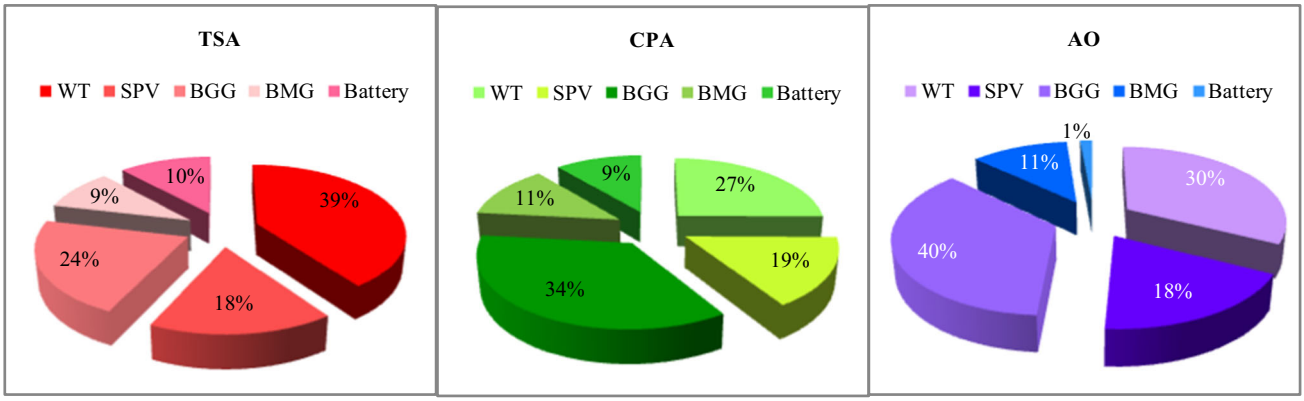


Fig. 14 Proportion of various components in the NPC

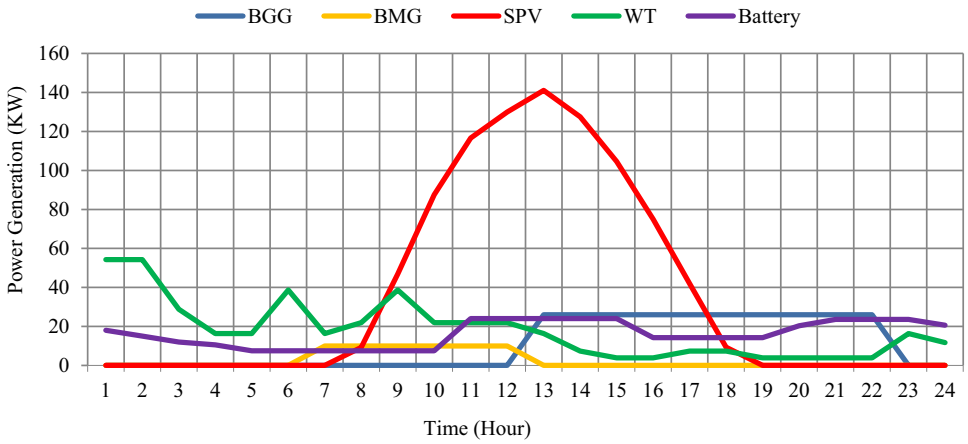


Fig. 15 Hourly Energy Generation for one day by different sources

Table 7 Statistical indicators for TSA, CPA and AO techniques

OT	Best	Average	Worst	SD	Median	Average RE	MAE	RMSE	Efficiency
TSA	649,787.5	669,837.5	689,850	25	689,852.412	0.0000867	1.6841625	12.00570	99.9927
CPA	659,812.5	670,850	690,475	31.25	689,854.712	0.0000732	1.9956337	24.89706	99.9913
AO	667,850	673,650	694,237.5	100	690,278.862	0.0011546	26.550912	134.7566	99.8849

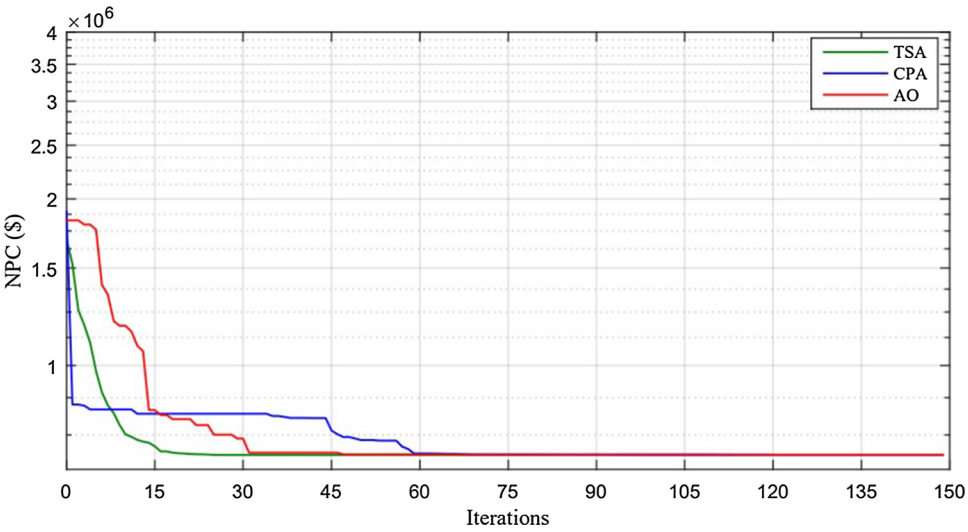


Fig. 16 Convergence curve for TSA, CPA and AO algorithms

- Further study can use other storage technologies, such as PHS and fuel cells.
- Moreover, hybrid methods combining metaheuristic and ML techniques can be developed to forecast RES.

Author contributions Professors M. Rizwan and P. Anand supervised and provided ideas for this research. Bandana Sharma has authored and compiled all of the study data used to reach at the conclusions. All of the authors read and the final manuscript was approved by all authors.

Declarations

Competing interests The authors declare no competing interests.

References

1. Sharma B, Rizwan M, Anand P (2023) A new intelligent approach for size optimization of a renewable energy based grid connected hybrid energy system. Int J Numer Model Electron Netw Devices Fields 36(2):1. https://doi.org/10.1002/jnm.3050
2. Markvart T, Fragaki A, Ross JN (2006) PV system sizing using observed time series of solar radiation. Sol Energy 80(1):46–50. https://doi.org/10.1016/j.solener.2005.08.011
3. Lujano-Rojas JM, Dufo-López R, Bernal-Agustín JL (2013) Probabilistic modelling and analysis of stand-alone hybrid power systems. Energy 63:19–27. https://doi.org/10.1016/j.energy.2013.10.003
4. Li J, Wei W, Xiang J (2012) A Simple Sizing Algorithm for Stand-Alone PV/Wind/Battery Hybrid Microgrids. Energies 5(12):5307–5323. https://doi.org/10.3390/en5125307
5. Zhang X, Tan S-C, Li G, Li J, Feng Z (2013) Components sizing of hybrid energy systems via the optimization of power dispatch simulations. Energy 52:165–172. https://doi.org/10.1016/j.energy.2013.01.013
6. Alturki FA, Al-Shamma'a AA, Farh HMH, AlSharabi K (2021) Optimal sizing of autonomous hybrid energy system using supply-demand-based optimization algorithm. Int J Energy Res 45(1):605–625. https://doi.org/10.1002/er.5766
7. Sadeghi D, Hesami Naghshbandy A, Bahramara S (2020) Optimal sizing of hybrid renewable energy systems in presence of electric vehicles using multi-objective particle swarm optimization. Energy 209:118471. https://doi.org/10.1016/j.energy.2020.118471
8. Mubaarak S et al (2021) Efficient photovoltaics-integrated hydrogen fuel cell-based hybrid system: Energy management and optimal configuration. J Renew Sustain Energy 13(1):013502. https://doi.org/10.1063/1.5141932
9. Li D, Zhu D, Wang R, Ge M, Wu S, Cai Y (2020) Sizing optimization and experimental verification of a hybrid generation water pumping system in a greenhouse. Math Probl Eng 2020:1–11. https://doi.org/10.1155/2020/3194196
10. Wu T, Zhang H, Shang L (2020) Optimal sizing of a grid-connected hybrid renewable energy systems considering hydroelectric storage. Energy Sources Part Recovery Util Environ Eff 1:1–17. https://doi.org/10.1080/15567036.2020.1731018
11. Das BK, Tushar MSHK, Hassan R (2021) Techno-economic optimisation of stand-alone hybrid renewable energy systems for concurrently meeting electric and heating demand. Sustain Cities Soc 68:102763. https://doi.org/10.1016/j.scs.2021.102763
12. Tabak A, Kayabasi E, Guneser MT, Ozkaymak M (2022) Grey wolf optimization for optimum sizing and controlling of a PV/WT/BM hybrid energy system considering TNPC, LPSP, and LCOE concepts. Energy Sources Part Recov Environ Eff 44(1):1508–1528. https://doi.org/10.1080/15567036.2019.1668880
13. Bernal-Agustín JL, Dufo-López R, Rivas-Ascaso DM (2006) Design of isolated hybrid systems minimizing costs and pollutant emissions. Renew Energy 31(14):2227–2244
14. Dufo-López R et al (2011) Multi-objective optimization minimizing cost and life cycle emissions of stand-alone PV–wind–diesel systems with batteries storage. Appl Energy 88(11):4033–4041
15. Yang H, Zhou W, Lu L, Fang Z (2008) Optimal sizing method for stand-alone hybrid solar–wind system with LPSP technology by using genetic algorithm. Sol Energy 82(4):354–367. https://doi.org/10.1016/j.solener.2007.08.005
16. Bilal BO, Sambou V, Ndiaye PA, Kébé CMF, Ndongo M (2010) Optimal design of a hybrid solar–wind-battery system using the minimization of the annualized cost system and the minimization of the loss of power supply probability (LPSP). Renew Energy 35(10):2388–2390
17. Tafreshi SMM, Zamani HA, Ezzati SM, Baghdadi M, Vahedi H (2010) Optimal unit sizing of distributed energy resources in microgrid using genetic algorithm. In: 2010 18th Iranian conference on electrical engineering, 2010, pp 836–841
18. Navaeefard A, Tafreshi SM, Barzegari M, Shahrood AJ (2010) Optimal sizing of distributed energy resources in microgrid considering wind energy uncertainty with respect to reliability. IEEE Int Energy Conf 2010:820–825
19. Bashir M, Sadeh J (2012) Optimal sizing of hybrid wind/photovoltaic/battery considering the uncertainty of wind and photovoltaic power using Monte Carlo. In: 2012 11th International conference on environment and electrical engineering, pp 1081–1086
20. Pirhaghshenasvali M, Asaei B (2014) Optimal modeling and sizing of a practical hybrid wind/PV/diesel generation system. In: The 5th Annual International Power Electronics, Drive Systems and Technologies Conference (PEDSTC 2014), pp 506–511
21. Askarzadeh A (2013) Developing a discrete harmony search algorithm for size optimization of wind–photovoltaic hybrid energy system. Sol Energy 98:190–195
22. Ismail MS, Moghavvemi M, Mahlia TMI (2014) Genetic algorithm based optimization on modeling and design of hybrid renewable energy systems. Energy Convers Manag 85:120–130
23. Maleki A, Askarzadeh A (2014) Optimal sizing of a PV/wind/diesel system with battery storage for electrification to an off-grid remote region: a case study of Rafsanjan, Iran. Sustain Energy Technol Assess 7:147–153
24. Heydari A, Askarzadeh A (2016) Optimization of a biomass-based photovoltaic power plant for an off-grid application subject to loss of power supply probability concept. Appl Energy 165:601–611
25. Abu-Hamdeh NH, Alnefaie KA (2019) Optimal selection and techno-economic analysis of a hybrid power generation system. J Renew Sustain Energy 11(5):055902
26. Diab AAZ, Sultan HM, Mohamed IS, Kuznetsov ON, Do TD (2019) Application of different optimization algorithms for optimal sizing of PV/wind/diesel/battery storage stand-alone hybrid microgrid. IEEE Access 7:119223–119245
27. Liu X, Li N, Mu H, Li M, Liu X (2021) Multiobjective optimization and parameters study of hybrid offshore wind and pumped hydrostorage power system: Modeling and case study. J Renew Sustain Energy 13(4):044103
28. Ganesan P, Xavier SAE (2022) Artificial cell swarm optimization and vapor liquid equilibrium for energy management system in smart grid. Int J Numer Model Electron Netw Devices Fields
29. Anand P, Bath SK, Rizwan M (2019) Renewable energy-based hybrid model for rural electrification. Int J Energy Technol Policy 15(1):86–113

30. Anand P, Bath SK, Rizwan M (2020) Size optimisation of RES-based grid connected hybrid power system using harmony search algorithm. Int J Energy Technol Policy 16(3):238–276
31. Saebi J, Seyyed Mahdavi S (2018) Optimal design of hybrid energy system based on persuasive policies for renewable resources in Iran (case study: University of Bojnord). J Renew Sustain Energy 10(4):045906
32. Rajesh P, Shajin FH, Rajani B, Sharma D (2022) An optimal hybrid control scheme to achieve power quality enhancement in micro grid connected system. Int J Numer Model Electron Netw Devices Fields, p e3019
33. Bakhshaei P, Askarzadeh A, Arababadi R (2021) Operation optimization of a grid-connected photovoltaic/pumped hydro storage considering demand response program by an improved crow search algorithm. J Energy Storage 44:103326
34. Jamshidi S, Pourhossein K, Asadi M (2021) Size estimation of wind/solar hybrid renewable energy systems without detailed wind and irradiation data: a feasibility study. Energy Convers Manag 234:113905
35. Rezaei M, Dampage U, Das BK, Nasif O, Borowski PF, Mohamed MA (2021) Investigating the impact of economic uncertainty on optimal sizing of grid-independent hybrid renewable energy systems. Processes 9(8):1468
36. Acharya S, Sivarajan G, Kumar DV, Srikrishna S (2021) A hybridized approach for design and optimization of combined economic emission dispatch. Energy Sources Part B Econ Plan Policy 16(10):903–928
37. Mishra S, Saini G, Chauhan A, Upadhyay S, Balakrishnan D (2023) Optimal sizing and assessment of grid-tied hybrid renewable energy system for electrification of rural site. Renew Energy Focus
38. Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z (2009) A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 13(4):915–920. https://doi.org/10.1016/j.rser.2008.02.002
39. Catalao JPS, Pousinho HMI, Mendes VMF (2010) Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Wind Power Forecasting in Portugal. IEEE Trans Sustain Energy, p 5571038. https://doi.org/10.1109/TSTE.2010.2076359.
40. Khalid M, Savkin AV (2010) A model predictive control approach to the problem of wind power smoothing with controlled battery storage. Renew Energy 35(7):1520–1526. https://doi.org/10.1016/j.renene.2009.11.030
41. Tascikaraoglu A, Uzunoglu M, Vural B (2012) The assessment of the contribution of short-term wind power predictions to the efficiency of stand-alone hybrid systems. Appl Energy 94:156–165. https://doi.org/10.1016/j.apenergy.2012.01.017
42. Hocaoglu FO, Gerek ON, Kurban M (2009) The effect of model generated solar radiation data usage in hybrid (wind–PV) sizing studies. Energy Convers Manag 50(12):2956–2963. https://doi.org/10.1016/j.enconman.2009.07.011
43. Zhang W, Maleki A, Rosen MA, Liu J (2019) Sizing a stand-alone solar-wind-hydrogen energy system using weather forecasting and a hybrid search optimization algorithm. Energy Convers Manag 180:609–621. https://doi.org/10.1016/j.enconman.2018.08.102
44. Gupta RA, Kumar R, Bansal AK (2015) BBO-based small autonomous hybrid power system optimization incorporating wind speed and solar radiation forecasting. Renew Sustain Energy Rev 41:1366–1375. https://doi.org/10.1016/j.rser.2014.09.017
45. Maleki A, Khajeh MG, Rosen MA (2016) Weather forecasting for optimization of a hybrid solar-wind–powered reverse osmosis water desalination system using a novel optimizer approach. Energy 114:1120–1134. https://doi.org/10.1016/j.energy.2016.06.134
46. Medina-Santana AA, Cárdenas-Barrón LE (2022) Optimal design of hybrid renewable energy systems considering weather forecasting using recurrent neural networks. Energies 15(23):9045. https://doi.org/10.3390/en15239045
47. Karimpour M, Schmid R, Tan Y (2021) Exact output regulation for wind turbine active power control. Control Eng Pract 114:104862
48. Kamjoo A, Maheri A, Putrus GA, Dizqah AM (2012) Optimal sizing of grid-connected hybrid wind-PV systems with battery bank storage. In: Proceedings of World Renewable Energy Forum
49. Tu J, Chen H, Wang M, Gandomi AH (2021) The colony predation algorithm. J Bionic Eng 18(3):674–710. https://doi.org/10.1007/s42235-021-0050-y
50. Abualigah L, Yousri D, Abd Elaziz M, Ewees AA, Al-qaness MAA, Gandomi AH (2021) Aquila Optimizer: A novel meta-heuristic optimization algorithm. Comput Ind Eng 157:1050. https://doi.org/10.1016/j.cie.2021.107250
51. Kaur S, Awasthi LK, Sangal AL, Dhiman G (2020) Tunicate Swarm Algorithm: A new bio-inspired based metaheuristic paradigm for global optimization. Eng Appl Artif Intell 90:103541

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

Optimal Shade Dispersion Strategy for Enhanced PV Power Extraction Under PSCs

Ranjeet Singh, Vinod Kumar Yadav, *Senior Member, IEEE* and Madhusudan Singh, *Senior Member, IEEE*

***Abstract*—This article proposes a static optimal shade dispersion (SD) strategy to alleviate the mismatch loss (ML) by uniformly dispersing the shade over the entire photovoltaic (PV) array. The proposed generalized module repositioning approach (GMRA) is valid for any $i \times j$ array dimensions, for $i = j$ or $i \neq j$, i and j being the number of rows and columns, respectively, and can assume even, odd, or both values. To extract optimal power from a series-parallel (SP) PV configuration under mismatch conditions, a variant of the proposed GMRA (SP-GMRA) is also presented. The module locations of the total-cross-tied (TCT) and SP-configured PV array are changed using GMRA and SP-GMRA respectively without altering the terminal electrical connections. Simulations are performed on a 9×9 PV array under four common shade occurrences, followed by experimental validation on symmetrical 5×5 and unsymmetrical 6×4 configurations to demonstrate the applicability and generalization of the proposed approach to varied array sizes. The proposed scheme is compared to a wide range of recently published state-of-the-art reconfiguration methods for performance evaluation both on simulation and hardware platforms. The proposed technique outperforms all other reconfiguration schemes and yields up to 76.3 % and 27 % lower ML compared to TCT and SP configurations, respectively.**

***Index Terms*--mismatch loss (ML), shade dispersion, partial shading conditions (PSCs), power generation, module relocation.**

I. Introduction

ATTAINMENT of energy independence depends mainly on the extent of power extraction from non-conventional energy sources. Among all available renewable energy sources, such as wind, hydro, solar, geothermal, and tidal, solar energy is considered as most promising in the world's progression toward clean and sustainable energy [1]. The wide acceptance of photovoltaic (PV) systems worldwide is attributed to the decrease in the manufacturing price of PV cells and recent developments in power transformation technology [2], [3]. However, one of the predominant issues faced by PV systems that very often impedes their optimal utilization is partial shading conditions (PSCs) [4], [5]. PV system experiences a decline in efficiency and performance because of PSCs [6],[7]. PSCs adversely affect PV array optimal utilization due to shade being cast upon by moving clouds, dust deposition, close communication lines trees, and houses nearby modules, thus causing a mismatch in generated current by various modules of the same array [8].

Under steady irradiance conditions (SICs), the presence of a distinct peak power point (PPP) makes it easier for the conventional PPP tracking algorithms to converge at the unique peak power operating point for optimum power extraction [9], [10]. Contrarily, under PSCs, uneven irradiance exposure to various array modules generates several peaks in power–voltage (P-V) characteristics. Subsequently, it causes mismatch losses (ML) [11]. Numerous approaches have been presented in the literature to access global peak (GP) under mismatch conditions [12]–[14]. The power extraction from the PV array under PSCs is a function of the shaded area, the position of shaded modules, shade patterns, and array interconnections.

Several conventional PV array interconnections, such as series-parallel (SP), honey-comb (HC), bridge link (BL), and total cross-tied (TCT), are utilized to reduce ML [15]. TCT arrangement shows improved response under PSC by restricting ML to a lower value than other configurations, thereby generating more output power. However, TCT suffers from non-uniform row current under mismatch conditions and causes multiple peaks in the power curve resulting in additional power loss [16]. The limitations of the conventional PV array configurations made it extremely important for researchers to look for new shade-resilient configurations for optimal PV power extraction. The electronic sensors and switches-based reconfiguration schemes, commonly referred to as electronic array reconfiguration (EAR) or dynamic reconfiguration approach (DRA), have also been reported [17], [18]. Here, electrical connections of the shade-affected modules within an array are dynamically reorganized to counter the shading effects. Although such reorganization approaches result in optimal shade dispersion (SD) in most cases, the complexity and cost involved in an extensive PV network limit their on-field deployment. The shortcomings of EAR approaches are overcome by static reconfiguration approaches having fixed electrical connections of the modules [19]–[25]. These approaches do not necessitate complex circuitry involving a network of sensors, switches, and intricate reconfiguration algorithms thus offering an economical option for efficient SD.

A twisted two-step repositioning approach (TTSA) was proposed for improved PV power extraction under PSCs for

Ranjeet Singh(email: ranjeetsingh_2k20phdee502@dtu.ac.in), Vinod Kumar Yadav (e-mail: vinodkumar@dtu.ac.in), and Madhusudan Singh (e-mail: madhusudan@dce.ac.in) are with the Centre of Excellence for Electric Vehicles and Related Technologies, Electrical Engineering Department, Delhi Technogical University Delhi-110042

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

9×9 PV array arrangement [26]. The rearrangement method allows the repositioning of PV modules to any location within an array. However, deviation from fixed column arrangement results in additional wiring/line losses as modules no longer remain serially connected. These losses are bound to increase manifold and dominate in extensive PV array networks. The image encrypted-based reconfiguration approach chaotic baker map (CBM) is not universally compatible with all array sizes (only valid for symmetrical array configurations) and hence severely limits its scalability [6]. Additionally, its SD is poor for odd symmetrical array sizes. Although applicable to all array sizes, the odd-even prime (OEP) [7] approach exhibits poor SD and multiple peak generations in the power curve. A Parquet SudoKu (PRS) approach was proposed for a 9×9 PV array [27]. The PRS technique is superior to many puzzle-based schemes because it disperses shade uniformly without repeating PV modules in adjacent subarrays. However, like other static reconfiguration techniques [8], [19], [20], [22], [23], [26], PRS has limited applicability for PV arrays of varying sizes, which makes it challenging for on-field deployment. Additionally, all of the aforementioned approaches, including PRS, reconfigured the modules initially connected in a TCT arrangement without considering other prominent array interconnections such as series-parallel (SP), which is the most widely employed connection nowadays along with TCT [15], [28]. In [15], the authors have proposed a different SD configuration for an SP connection. However, the reported technique failed to generate more power beyond that of the SP connection.

To address the limitations of the aforementioned approaches, a novel static optimal SD strategy designated as generalized module repositioning approach (GMRA) is proposed for enhanced power extraction under PSCs. Unlike previously reported methods, to extract enhanced power output from an SP-configured PV array under PSCs, a variant of the proposed algorithm designated as SP-GMRA is also suggested in this article. Another significant contribution of the proposed work, missing in the majority of the previously reported works, is the universal applicability of the proposed approach to any *i*×*j* PV array, where *i* and *j* represent the number of rows and columns respectively, for *i*=*j* or $i \neq j$ where *i, j* can take even, odd, or both values. At the time of installation, the modules of the traditional TCT and SP arrangements are repositioned only once, with the electrical connections remaining unchanged. The proposed reconfiguration strategy is the improved alternative for enhanced output yield in a typical 9×9 array arrangement over benchmark puzzle-based rearrangement strategies. The applicability of the proposed approach is initially tested through simulation utilizing MATLAB/Simulink on a 9×9 PV array and later experimentally validated on an outdoor hardware setup comprising 25 PV modules interconnected in a symmetrical 5×5 and unsymmetrical 6×4 PV array arrangements. Different ratings of modules are selected for simulation and experimental analysis to demonstrate the application of GMRA and SP-GMRA across varying power levels. Four different commonly occurring shading scenarios: bottom right corner (BRC), top right corner (TRC), top left corner (TLC), and bottom left corner (BLC) are selected to extensively investigate the performance of the proposed approach on a PV array exposed to shading on different sections. Artificial shading scenarios are generated using different layers of transparent plastic sheets imitating intense and moderate shading events. Recent state-of-the-art reconfiguration methods are selected for simulation and experimental validation. This allows for the performance comparison of the proposed SD strategy with recent reconfiguration methods. The proposed reconfiguration strategy is simple and does not require any complex network of switches and sensors, even for large-sized PV installations.

The remainder of the paper is structured as follows: the upcoming section elucidates the electrical behavior of a PV array under various operating conditions, followed by the description of different PV array configurations in section III. The simulation and experimental validation of the proposed GMRA and SP-GMRA is reported in section IV, followed by efficiency estimation and techno-economic analysis in section V. The conclusions are finally drawn in section VI.

II. Electrical Characteristics of a PV Array

Understanding the electrical behavior of PV arrays under various operating scenarios and the underlying principles of optimal SD is essential for developing an efficient SD strategy for maximizing power harvesting from a PV array. The PV array displays intriguing behavior as a result of high correlations with numerous external parameters, with temperature and irradiance standing out as the most significant. The modeling of a PV array and its response is considered next.

A. Steady Irradiance Condition (SIC)

Several PV cells connected serially make up a PV module. The single-diode model shown in Fig. 1(a) is employed in this work [20]. The voltage (V_G) and generated current (I_G) of a PV module are related:

$$I_G = I_L - I_{rev}\left[exp\left(\frac{V_G + R_s I_G}{kV_{Th}}\right) - 1\right] - \frac{V_G + R_s I_G}{R_p} \tag{1}$$

where I_L is the module's light-induced current, I_{rev} is the reverse saturation current of the electrical diode equivalent, k is ideality factor, V_{Th}, represents thermal voltage. Further, R_s and R_p represents module equivalent series, and shunt resistances respectively. Since the light-induced current, I_L in (1) is dependent on irradiance, S_a and temperature, T [21], the panel-generated current I_G is also expressed as

$$I_G(S_a, T) = I_{SC,0}\left(\frac{S_a}{S_{as}}\right)\left[1 + \gamma\left(T - T_s\right)\right]\frac{\left(R_s + R_p\right)}{R_p} \tag{2a}$$

where $I_{SC,0}$ is the module's current corresponding to the short circuit at standard irradiance and temperature conditions $\left(S_{as} = 1000\, W/m^2;\, T_s = 25\,^0C\right)$ and γ represents the temperature-dependent constant of the module's current.

$$I_G = I_{SC,0} \times \left(\frac{S_a}{S_{as}}\right) = \alpha \times I_{SC,0} \quad \forall T = T_s \tag{2b}$$

where the shading ratio, $\alpha = \left(\frac{S_a}{S_{as}}\right), \alpha \in [0,1]$ where $\alpha = 1$ and $\alpha = 0$ corresponds to a fully illuminated and completely shaded

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

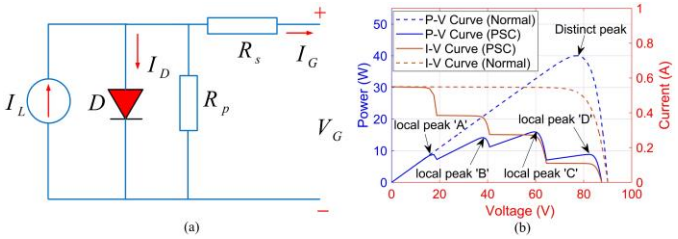


Fig. 1. (a) Electrical circuit model of a PV cell (b) PV electrical characteristics under SIC and PSCs (multiple peak generation).

module respectively. The I-V characteristic equation under SIC for any arbitrary-sized array comprising N_s series modules in each N_p shunt connections is as follows:

$$I_{pv} = N_p \times \left(I_L - I_{rev}\right)\left[exp\left(\frac{V_G + \frac{N_s}{N_p} R_s I_G}{k N_s V_{Th}} \right) - 1 \right] + \frac{V_G + \frac{N_s}{N_p} R_s I_G}{\frac{N_s}{N_p} R_p} \quad (3)$$

Using equations (1), (2), and (3), the PV module and array considered in this work are modeled.

B. Under PSCs

The bypass diode (BPD) gets activated due to radiation mismatch among modules under PSCs. BPDs provide an alternate path for the current to flow, thus bypassing the shaded module and averting its possible damage due to hotspot formation, though not completely. This results in the formation staircase I-V curve with the appearance of multiple peaks in the resultant P-V curve [5]. Fig. 1(b) shows typical electrical characteristics of a PV array under SIC and PSC.

C. Estimation of ML for Optimal SD

For any $i \times j$ PV array arrangement, to quantify the effectiveness of any SD strategy, estimation of ML is vital and can be obtained from the difference between the resultant sum of individual maximum generated module powers and the total array power under PSCs as follows:

$$ML = \sum_{q=1}^{i \times j} P_{m,q} - P_{array} \quad (4)$$

where $P_{m,q}$ and P_{array} are, respectively, the optimal generated powers from the q^{th} module and the entire PV array under PSCs. Fig. 2 (a) shows a typical 2×2 shading event on a 4×4 PV array with 4 modules under the impact of shading. The array can be partitioned into four individual groups connected serially, each having four parallel PV units. Here, the current generated by the unshaded and shaded modules is I_G and αI_G respectively. As mentioned earlier, 'α' is the shading ratio. The optimal SD achieved after module reconfiguration is shown in Fig. 2(b). The ML is estimated under two cases:

1) Case I- 4×4 TCT array under 2×2 shading (Prior SD):

Two modules each in groups III and IV are shaded. The 12 unshaded modules and 4 shaded modules individually generate maximum powers of $(12V_G \times I_G)$ and $4\alpha I_G \times V_G$ respectively. For the estimation of the global peak power of the PV array, P_{array}, first, the currents of the individual groups of the array are determined. Since different groups contribute different currents, the existence of multiple peaks in the PV curve is inevitable. The panels in group IV will be bypassed first, followed by group-III, II, and finally group-I having voltages of $4V_G$, $3V_G$, $2V_G$, and V_G respectively. The array peak power under this condition is estimated to be $8V_G \times I_G(1+\alpha)$. The ML using (4) can be estimated as:

$$ML\big|_{4\times4}^{prior\ SD} = \underbrace{\left[(12V_G \times I_G) + 4\alpha I_G \times V_G\right]}_{\sum_{q=1}^{4\times4} P_{m,q} \quad \forall\ q=1,2,3...16} - \underbrace{\left[8V_G \times I_G(1+\alpha)\right]}_{P_{array}} \quad (5a)$$

$$= 4(1-\alpha)V_G \times I_G$$

2) Case II- 4×4 PV array modules reconfigured to obtain optimal SD using a repositioning approach (Post SD):

Post application of the SD approach, the shade gets dispersed across various rows of the PV array, thus minimizing the group current difference. Each group now has only one shaded module. The resultant ML can be estimated as before:

$$ML\big|_{4\times4}^{post\ SD} = \underbrace{\left[(12V_G \times I_G) + 4\alpha I_G \times V_G\right]}_{\sum_{q=1}^{4\times4} P_{m,q} \quad \forall\ q=1,2,3...16} - \underbrace{\left[4V_G \times (3+\alpha) \times I_G\right]}_{P_{array}}$$

$$= 0 \quad (5b)$$

In this manner, an effective SD strategy disperses any shade pattern to alleviate the ML, thus ensuring the optimal utilization of the PV array arrangement. Fig. 2(c) shows the variation in ML as a function of shading ratio, α, prior and post SD. It can be observed that post-SD, the ML decreases considerably.

III. Description Of PV Array Configurations

A. PV Array in TCT Configuration

The power extraction from a PV array configured in a TCT arrangement is analyzed under various PSCs. The output voltage and current extracted from a $i \times j$ TCT-arranged PV array can be expressed as given by

$$I_{pv} = \sum_{q=1}^{j} I_{G,pq} \quad \forall\ p = 1, 2, ..., (i-1) \quad (6)$$

$$V_{pv} = \sum_{p=1}^{i} V_{G,pq} \quad \forall\ q = 1, 2, ..., (j-1) \quad (7)$$

The following equation gives the two representations of the PV array used in this work:

$$M_{i\times j} = \underbrace{\begin{matrix} C_1 & C_2 & C_3 & C_j \\ \overbrace{\quad\quad\quad\quad}^{Columns} & & & \end{matrix}}_{} \underbrace{\begin{bmatrix} 11 & 12 & 13\cdots1j \\ 21 & 22 & 23\cdots2j \\ \vdots & \vdots & \vdots\cdots\vdots \\ i1 & i2 & i3\cdots ij \end{bmatrix}}_{Form-I\ \ i\times j} = \underbrace{\begin{matrix} C_1 & C_2 & C_3 & C_j \\ \overbrace{\quad\quad\quad\quad}^{Columns} & & & \end{matrix}}_{} \underbrace{\begin{bmatrix} 1 & 1 & 1\cdots1 \\ 2 & 2 & 2\cdots2 \\ \vdots & \vdots & \vdots\cdots\vdots \\ i & i & i\cdots i \end{bmatrix}}_{Form-II\ \ i\times j} \quad (8)$$

The module parametric specifications at standard test condition (STC) employed for simulation and experimental analysis are listed in Table I.

B. Proposed GMRA and SP-GMRA for Optimal SD

In any PV array arrangement, each PV module position can be identified by a definite row and column index value as depicted in (8). For instance, module position '$1j$' corresponds

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

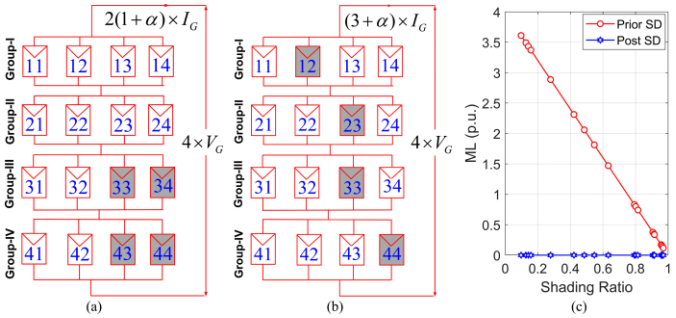


Fig. 2. (a) 4×4 TCT under a typical 2×2 PSC (b) optimal SD through a repositioning approach (c) ML as a function of SR prior and post application of SD approach.

TABLE I
PARAMETRIC DATA OF A PV MODULE AT STC USED IN THIS WORK

Parameter	Simulation	Experimental
Optimal power output (P_{max})	50 W	10 W
Voltage at MPP ($V_{pv,m}$)	19.95 V	20 V
Current at MPP ($I_{pv,m}$)	2.51 A	0.50 A
Short circuit current (I_{sc})	2.63 A	0.60 A
Open circuit Voltage (V_{oc})	22.26 V	24.8 V
Dimensions (L×W×H)	430×665×35 mm	255×350×22 mm

to a row index value '1' and column index value ' j '. Only the row indices are modified in the proposed GMRA, while the column indices remain unaltered.

1) Case I: $i \times j$ PV array, where, i= odd number:

Step 1: Identify the primary column, C_k , in the PV array, where $k = \left\lceil \frac{j}{2} \right\rceil$, the ceil operator '$\lceil\ \rceil$', rounds a real number to the lowest next integer. Next, arrange the primary column row index values sequentially in ascending order as shown in (9). Steps 2 and 3 describe the further restructuring of the PV array located to the right and left of the primary column, respectively.

$$M_{r,i\times j} = \begin{array}{c} \begin{matrix} C_1 & \cdots & C_{k-1} & C_k & C_{k+1} & C_{k+2} & \cdots & C_j \end{matrix} \\ \begin{bmatrix} \frac{3}{2}m-1 & \cdots & m-1 & 1 & m & 2 & \cdots & \frac{m}{2} \\ \frac{3}{2}m & \cdots & m & 2 & m+1 & 3 & \cdots & \frac{m+2}{2} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \frac{m+2}{2} & \cdots & 1 & m & 2 & m+1 & \cdots & \frac{3}{2}m-1 \\ \frac{m}{2}+2 & \cdots & 2 & m+1 & 3 & m+2 & \cdots & \frac{3}{2}m \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \frac{3}{2}m-2 & \cdots & m-2 & j & m-1 & 1 & \cdots & \frac{m-2}{2} \end{bmatrix}_{i\times j} \end{array} \quad (9)$$

Step 2: Divide the entire PV array located to the right of the primary column, C_k , including the primary column itself, into two groups i.e. top and bottom w.r.t row index values. The bottom group starts with the m^{th} row index of the primary column, where $m = \left\lceil \frac{i}{2} \right\rceil + 1$ and continues sequentially until the last row index in the column. The remaining row index values of the primary column constitute the top group. The row index values in the subsequent column, C_{k+1} are obtained by replacing the top and bottom group row index values of the column with the bottom and top group row index values of the preceding column, C_k , respectively in a sequential order and so on.

Step 3: Divide the entire PV array located to the left of the primary column, C_k , including the primary column itself, into two groups i.e. top and bottom w.r.t row index values. The bottom group starts with the $(m-1)^{th}$ row index value of the primary column, C_k , and continues sequentially until the last row index in the column. The remaining row index values of the primary column constitute the top group. The row index values in the preceding column, C_{k-1} are achieved by replacing the top and bottom group row index values of the column with the bottom and top group row index values of the subsequent column, C_k , respectively in a sequential manner and so on.

Likewise, the row index values of the remaining columns are obtained following the aforementioned steps and the final reconfigured TCT array using GMRA is obtained. After the application of GMRA, a 9×9 PV array is shown in (10).

$$M_{r,9\times9} = \begin{array}{c} \text{Columns} \\ \begin{matrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 \end{matrix} \\ \begin{bmatrix} 81 & 42 & 93 & 54 & 15 & 66 & 27 & 78 & 39 \\ 91 & 52 & 13 & 64 & 25 & 76 & 37 & 88 & 49 \\ 11 & 62 & 23 & 74 & 35 & 86 & 47 & 98 & 59 \\ 21 & 72 & 33 & 84 & 45 & 96 & 57 & 18 & 69 \\ 31 & 82 & 43 & 94 & 55 & 16 & 67 & 28 & 79 \\ 41 & 92 & 53 & 14 & 65 & 26 & 77 & 38 & 89 \\ 51 & 12 & 63 & 24 & 75 & 36 & 87 & 48 & 99 \\ 61 & 22 & 73 & 34 & 85 & 46 & 97 & 58 & 19 \\ 71 & 32 & 83 & 44 & 95 & 56 & 17 & 68 & 29 \end{bmatrix}_{9\times9} \end{array} \quad (10)$$

Step 4: To obtain the final reconfigured SP PV array using SP-GMRA, the row and column index values obtained using GMRA are interchanged before implementation on an SP configuration. This means that the module positions identified as '81', '45', and '36' in GMRA marked by dotted circles in (10) correspond to '18', '54', and '63' respectively in SP-GMRA, and so on.

2) Case II: $i \times j$ PV array, where i= even number:

Step 1: Identify the primary column, C_k , in the PV array, where $k = \frac{j}{2} + 1$ *for* $j =$ *even number* , and $k = \left\lceil \frac{j}{2} \right\rceil$ *for* $j =$ *odd number* . Next, sequentially arrange, C_k , row indices in ascending order.

Step 2: Divide the entire PV array into two halves (groups) including the primary column, C_k , w.r.t row indices.

Step 3(a): The module positions for the subsequent column, C_{k+1} is achieved by replacing the top and bottom group row index values in the column with the bottom and top group row indices of the preceding column, C_k , respectively, and so on.

Step 3(b): The module positions for C_{k+2} are achieved by first shifting the row index values in an anticlockwise manner, beginning with the bottom-most member for each group of the preceding column C_{k+1}, followed by the replacement of the top and bottom group row index values of the column with bottom and top group row index values of the preceding column, C_{k+1} ,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

respectively. This is done to avoid any repetition of the columns. The same procedure is applied for the remaining columns positioned to the right.

***Step 4(a)*:** The row indices for C_{k-1} and preceding columns except C_1 is achieved by following the same procedure as described above in step 3(b).

***Step 4(b)*:** The first column, C_1 row indices are obtained by performing a similar operation on the column C_2 as described in step 3(a). This step is followed for all array configurations of the form $i \times j$, where $i = j$ or $i \neq j$. An unsymmetrical 6×4 PV array is repositioned and represented in (11) post-GMRA.

$$M_{r,6\times4} = \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} \\ \left[\begin{array}{cccc} 3\,\text{C} & 6\,\text{C} & 1\,\text{A} & 4\,\text{A} \\ 1\,\text{A} & 4\,\text{A} & 2\,\text{B} & 5\,\text{B} \\ 2\,\text{B} & 5\,\text{B} & 3\,\text{C} & 6\,\text{C} \\ 6\,\text{C} & 3\,\text{C} & 4\,\text{A} & 1\,\text{A} \\ 4\,\text{A} & 1\,\text{A} & 5\,\text{B} & 2\,\text{B} \\ 5\,\text{B} & 2\,\text{B} & 6\,\text{C} & 3\,\text{C} \end{array}\right]_{6\times4} \end{array} \tag{11}$$

The circuit arrangement of the TCT configuration and proposed GMRA for a 5×5 PV array with fixed electrical connections is shown in Fig. 3(a) and (b) respectively.

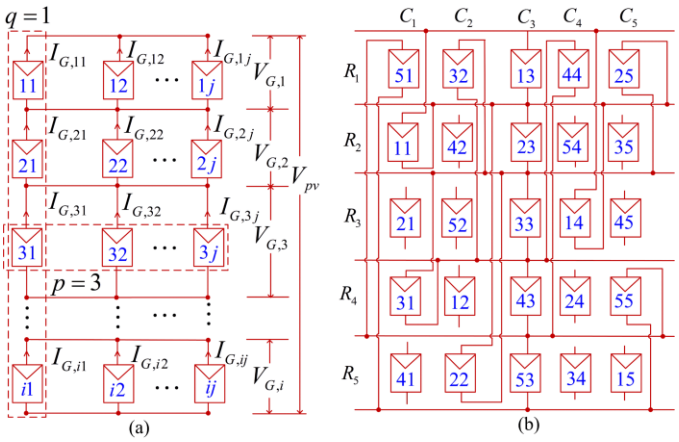


Fig. 3. (a) i×j PV array connected in TCT configuration (b) GMRA with the fixed electrical connection of PV modules for a 5×5 PV array.

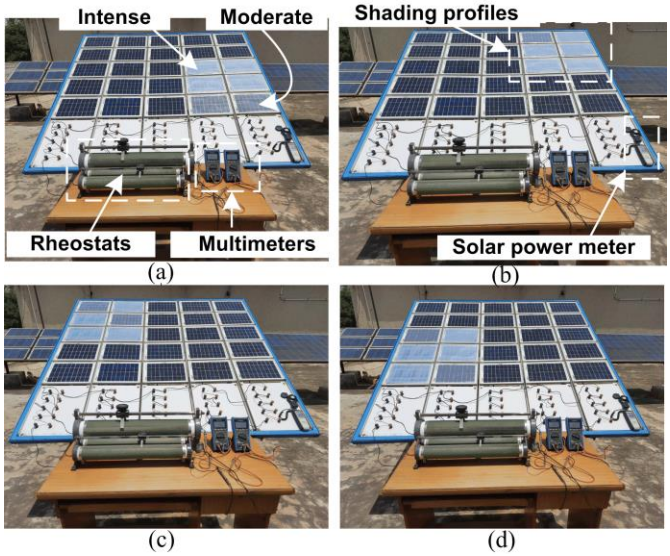


Fig. 4. Experimental prototype (5×5 PV array) for hardware validation of the performance of the proposed technique under different shading cases (a) BRC (b) TRC (c) TLC (d) BLC with 'intense' and 'moderate' shading profiles.

IV. Simulation Results and Hardware Implementation

To validate the efficacy and robustness of the proposed GMRA and SP-GMRA, PV arrays configured in the two most prominently employed topologies viz. TCT and SP are examined. A scaled-down experimental setup was established, as shown in Fig. 4, for hardware validation under different shading scenarios. The experimental arrangement comprises 25 identical PV modules, multimeters, rheostats, and a solar power meter. Simulations are performed on a 9×9 PV array, followed by experimental validation on symmetrical 5×5 and unsymmetrical 6×4 configurations to demonstrate the applicability and generalization of the proposed approach to varied array sizes.

A. Simulation and Performance Analysis

The proposed GMRA's performance is evaluated against the benchmark TCT configuration and three recently published state-of-the-art reconfiguration methods i.e. CMAP [6], PRS [27], and TTSA [26] under four different shading cases i.e. BRC, TRC, TLC, and BLC. Fig. 5 and 6 respectively show the SD performance and power voltage characteristics of a 9×9 PV array with different reconfiguration approaches under the impact of aforementioned shading scenarios. The comparative performance evaluation is done utilizing four important performance metrics such as fill factor (FF), mismatch loss (ML), execution ratio (ER), and power enhancement. Additionally, the performance metrics of SP-GMRA are also evaluated against the commonly employed SP configuration. The power enhancement is assessed against TCT and SP configurations for GMRA and SP-GMRA respectively. Table II gives a comparative performance estimation of the proposed approach in terms of the aforementioned performance metrics against TCT, CMAP, PRS, TTSA, and SP. To comprehensively examine the overall performance of the proposed technique under different shading scenarios, the average values of various performance metrics from Table II are considered.

1) Fill factor (FF): It is a measure of the squareness of the I-V characteristics under PSCs, hence an extremely important metric for quantifying the effectiveness of any SD approach. The proposed GMRA produced on average, 12.76 %, 12.76 %, 3.87 %, and 2.14 % higher FF than TCT, CMAP, PRS, and TTSA arrangements respectively under all considered shade cases. Additionally, on reconfiguring the traditional SP configuration using the proposed SP-GMRA, on average 7.1 % higher FF was achieved.

2) Mismatch Losses (ML): The ML is estimated similarly as described earlier in (5a) and (5b). On taking the average of all ML values under different shading cases, the proposed GMRA generates 76.3 %, 76.2 %, 51.2 %, and 37 % lower ML compared to TCT, CMAP, PRS, and TTSA configurations respectively. Moreover, the SP-GMRA arrangement generates 27 % lower ML than the SP array.

3) Execution Ratio (ER): The average ER obtained with TCT, CMAP, PRS, and TTSA is 76.5 %, 76.5 %, 83.0 %, and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

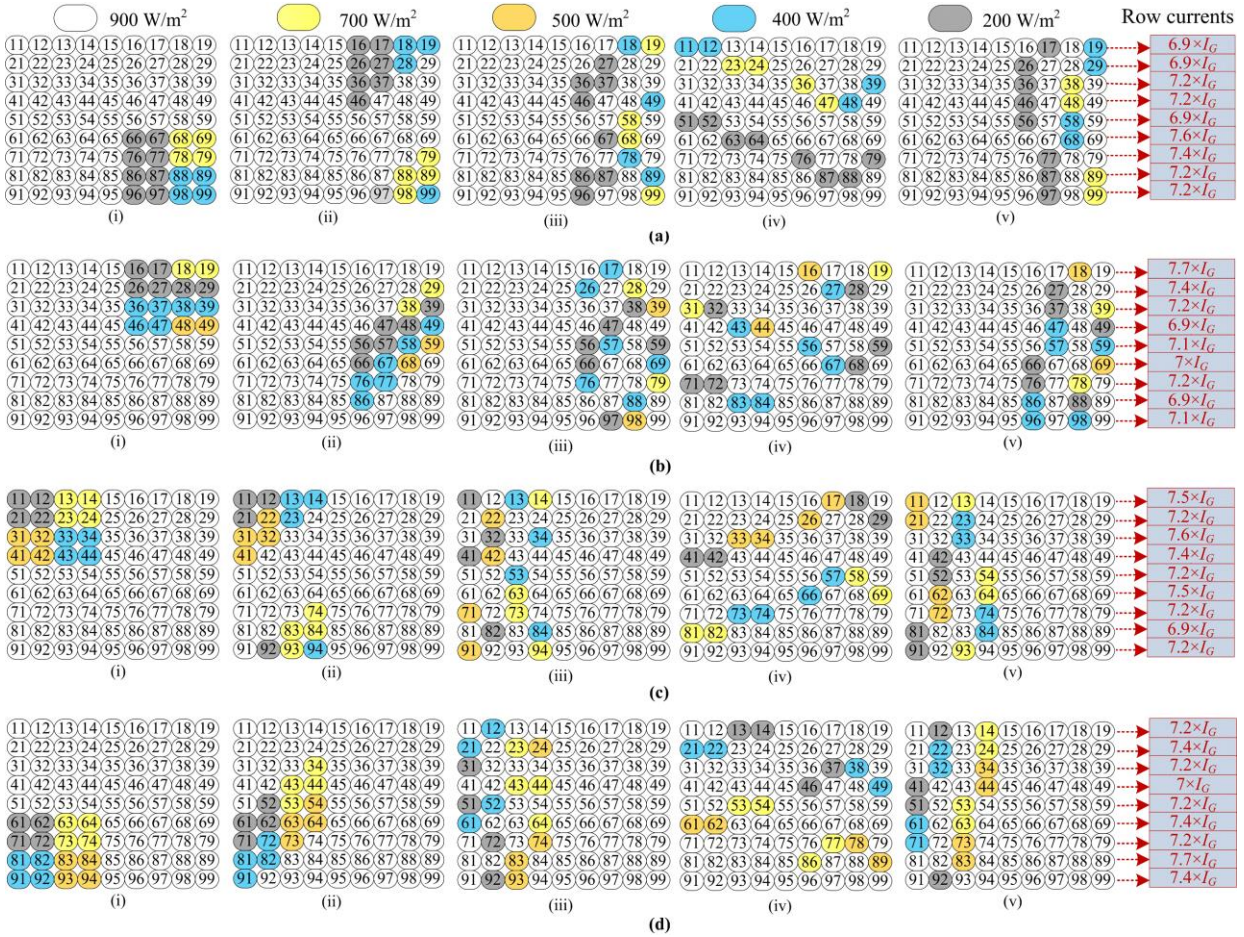


Fig. 5. SD performance on a 9×9 PV array under (a) BRC shading scenario (b) TRC shading scenario (c) TLC shading scenario (d) BLC shading scenario with (i) TCT arrangement (ii) CMAP [6] arrangement (iii) PRS [27] arrangement (iv) TTSA [26] arrangement (v) proposed GMRA.

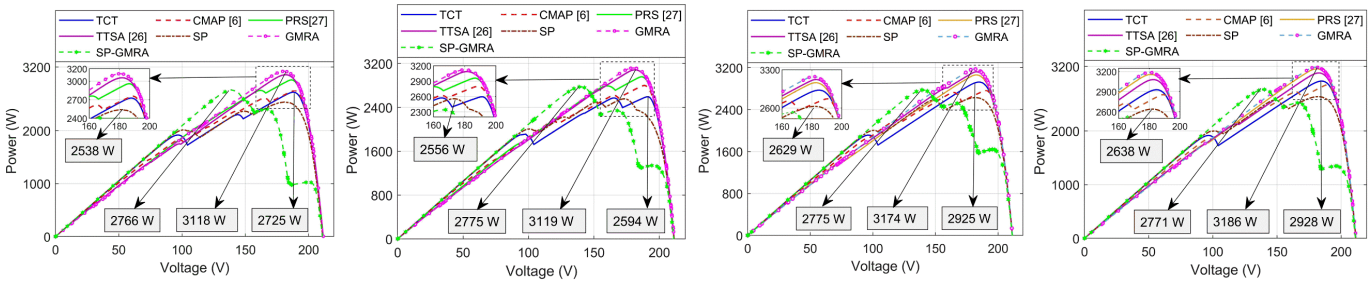


Fig. 6. Simulated power-voltage characteristics of 9×9 PV array under different shading scenarios (a) BRC (b) TRC (c) TLC (d) BLC.

84.48 % respectively. On the contrary, the proposed GMRA exhibits an average ER of 86.2 %. The SP-GMRA offers ER on average, 7% higher than traditional SP configuration.

4) Power Enhancement: The proportionate power enhancement w.r.t TCT is used as a measure to evaluate the relative performance of the proposed GMRA method with CMAP, PRS, and TTSA, while for the SP-GMRA, the performance enhancement is evaluated w.r.t to SP configuration as shown in Table II. On average, the power enhancement attained w.r.t TCT configuration using CMAP, PRS, TTSA and proposed GMRA is 0.03 %, 9.02 %, 10.41%, and 12.75% respectively. Additionally, the power enhancement of 7% is achieved when SP-GMRA arrangement is used over SP connection under different shading conditions. In addition to SD, the row current computations are also shown for the proposed scheme (refer Fig. 5). The proposed scheme produces uniform row (group) currents throughout the array and consistently exhibits the best SD in all cases. As evident from Table II, the proposed scheme outperforms all other reconfiguration methods on prominent performance metrics. The simulated power-voltage characteristics under different shading conditions are shown in Fig. 6. The global peak powers attained with TCT, SP, GMRA, and SP-GMRA configurations are also highlighted. The TCT and SP array generates the highest global peak powers when the PV arrays are reconfigured using GMRA and SP-GMRA respectively. Among all considered reconfiguration approaches, the proposed technique exhibits clear dominance.

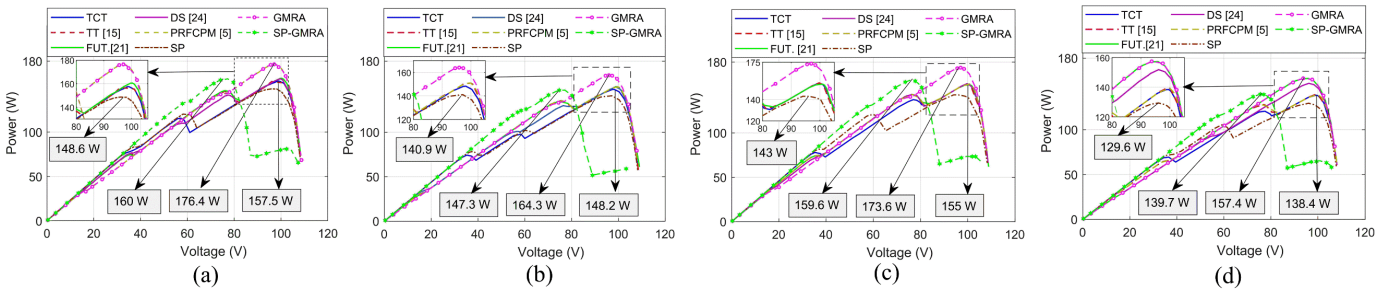
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

TABLE II
COMPARATIVE PERFORMANCE ANALYSIS OF THE PROPOSED TECHNIQUE WITH THE EXISTING STATE OF THE ART RECONFIGURATION APPROACHES

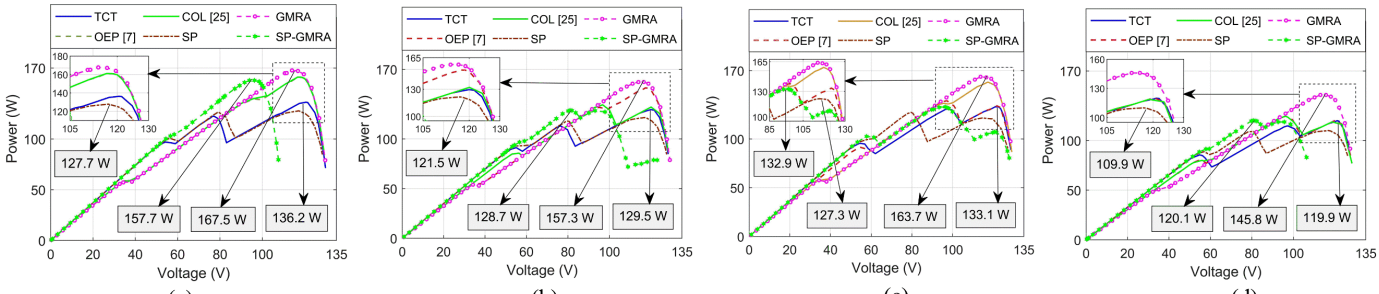
Shading scenarios	PV configuration	I_{mp}	V_{mp}	P_{mp}	Parametric indicators/Metrics				Optimal configuration
					Fill factor (%)	Mismatch loss (W)	Execution ratio (%)	Power Enhancement (%)	
BRC	TCT	14.49	188.05	2725	63.8	505	74.6	N.A	*GMRA*
	CMAP [6]	14.56	189.3	2756	64.5	474	75.5	1.13	
	PRS [27]	15.84	186.64	2956	69.2	274	80.9	8.47	
	TTSA [26]	16.79	181.77	3052	71.5	178	83.6	12	
	SP	14.04	180.76	2538	59.4	692	69.5	N.A	
	GMRA	17.24	180.88	**3118**	**73**	**112**	**85.4**	**14.42**	
	SP-GMRA	19.93	138.8	**2766**	**64.8**	**464**	**75.7**	**8.98**	
TRC	TCT	13.53	191.64	2594	60.7	636	71.0	N.A	*GMRA*
	CMAP [6]	14.81	188.97	2799	65.5	431	76.6	7.9	
	PRS [27]	15.84	186.77	2959	69.3	271	81.0	14.07	
	TTSA [26]	17.05	180.89	3084	72.2	146	84.5	18.88	
	SP	14.74	173.39	2556	59.9	674	70.0	N.A	
	GMRA	17.36	179.64	**3119**	**73.1**	**111**	**85.4**	**20.24**	
	SP-GMRA	19.81	140.1	**2775**	**65.0**	**455**	**76**	**8.56**	
TLC	TCT	15.78	185.38	2925	68.5	365	80.2	N.A	*GMRA*
	CMAP [6]	14.58	189.75	2766	64.8	524	75.77	-5.43	
	PRS [27]	16.82	181.99	3062	71.7	228	83.8	4.68	
	TTSA [26]	17.05	182.96	3122	73.1	168	85.52	6.73	
	SP	14.39	182.63	2629	61.6	661	72.0	N.A	
	GMRA	17.56	180.73	**3174**	**74.3**	**116**	**86.9**	**8.51**	
	SP-GMRA	19.93	139.18	**2774**	**64.9**	**516**	**76**	**5.51**	
BLC	TCT	15.84	184.7	2928	68.6	362	80.2	N.A	*GMRA*
	CMAP [6]	15.09	189.2	2855	66.9	435	78.2	-2.49	
	PRS [27]	17.37	181.67	3155	73.9	135	86.4	7.75	
	TTSA [26]	16.91	182.03	3078	72.1	212	84.3	5.12	
	SP	14.53	181.55	2638	61.8	652	72.2	N.A	
	GMRA	17.68	180.18	**3186**	**74.6**	**104**	**87.2**	**8.81**	
	SP-GMRA	19.77	140.14	**2771**	**64.9**	**519**	**75.9**	**5.04**	

The bold entities indicate that the proposed GMRA and SP-GMRA yielded outstanding results under different conditions.



(a) (b) (c) (d)

Fig. 7. Experimental power-voltage characteristics of 5×5 PV array for different shading scenarios (a) BRC shading scenario (b) TRC shading scenario (c) TLC shading scenario (d) BLC shading scenario.



(a) (b) (c) (d)

Fig. 8. Experimental power-voltage characteristics of 6×4 PV array for different shading scenarios (a) BRC shading scenario (b) TRC shading scenario (c) TLC shading scenario (d) BLC shading scenario.

B. Hardware Implementation

Commercial-grade transparent plastic sheets (200 microns) are utilized to generate artificial shading scenarios. Multiple layers of sheets are employed to create 'intense' and 'moderate' shading profiles (refer Fig. 4). 25 PV modules of 10 W each are configured in symmetrical 5×5 and unsymmetrical 6×4 arrangements. Four common shading profiles were created i.e. BRC, TRC, TLC, and BLC on the two considered PV arrangements configured in TCT and SP configurations. The hardware validation of the performance of the proposed GMRA

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

is performed against many popular and recently published state-of-the-art methods i.e. PRFCM [5], OEP [7], TT [15], FUT [21], DS [24], COL [25]. Fig. 7 and 8 shows the experimental power-voltage characteristics of 5×5 and 6×4 PV array under predefined shading conditions. SP-GMRA configuration is tested against conventional and more common SP configuration. Under all shading conditions, the proposed approaches, GMRA (reconfigured TCT) and SP-GMRA (reconfigured SP) generated the highest peak power points against all other reconfiguration methods and SP configuration respectively as evident from Fig. 7 and 8. The average power enhancement of 29.55 % and 22.3 % respectively is attained using the proposed GMRA against TCT arrangement under all considered shading conditions for 5×5 and 6×4 PV arrays. Similarly, compared to the SP configurations, 7.93 %, and 10.9 % higher peak power is extracted using SP-GMRA when tested on 5×5 and 6×4 arrangements respectively.

V. Efficiency Computations and Techno-economic Feasibility Study

A. Calculation of Efficiency

The accurate assessment of the efficiency plays a crucial role in effectively comparing different reconfiguration schemes for the PV system under PSCs. For any $i \times j$ PV array arrangement, the efficiency, $\eta|_x^y$, can be estimated as depicted in (12).

$$\eta|_x^y = \frac{GMPP|_x^y}{\left(\sum_{m=1}^{i}\sum_{n=1}^{j} S_{mn}\right) \times A_p} \quad \begin{matrix} \forall\ m = 1,\ 2,\ 3,\ \ldots,\ i \\ \forall\ n = 1,\ 2,\ 3,\ \ldots,\ j \end{matrix} \qquad (12)$$

where 'x' and 'y' denotes the shading scenario and PV configuration respectively. Additionally, '$S_{mn}(W/m^2)$' represents the irradiance input received by the module located at m^{th} row and n^{th} column and $A_p(m^2)$ denotes the area of the individual module, which is 0.285 m^2 (from Table I). Using (12), the efficiency calculations for TCT, CMAP [6], PRS [27], TTSA [26], SP, GMRA, and SP-GMRA under BRC, TRC, TLC, and BLC shading conditions (see Fig. 5), are performed. Fig. 9 illustrates the comparative performance in terms of system efficiency yields for a 9×9 PV array utilizing different

TABLE III
Efficiency of 9×9 PV Array Under Different Shading Scenarios

Shading scenario	PV configuration	Efficiency (%)	Shading scenario	PV configuration	Efficiency (%)
BRC	TCT	14.77	TRC	TCT	14.06
	CMAP [6]	14.94		CMAP [6]	15.17
	PRS [27]	16.02		PRS [27]	16.04
	TTSA [26]	16.54		TTSA [26]	16.72
	SP	13.76		SP	13.85
	GMRA	**16.9**		GMRA	**16.91**
	SP-GMRA	14.99		SP-GMRA	15.04
TLC	TCT	15.57	BLC	TCT	15.87
	CMAP [6]	14.89		CMAP [6]	15.48
	PRS [27]	16.29		PRS [27]	17.1
	TTSA [26]	16.61		TTSA [26]	16.68
	SP	13.99		SP	14.04
	GMRA	**16.89**		GMRA	**16.95**
	SP-GMRA	14.76		SP-GMRA	14.75

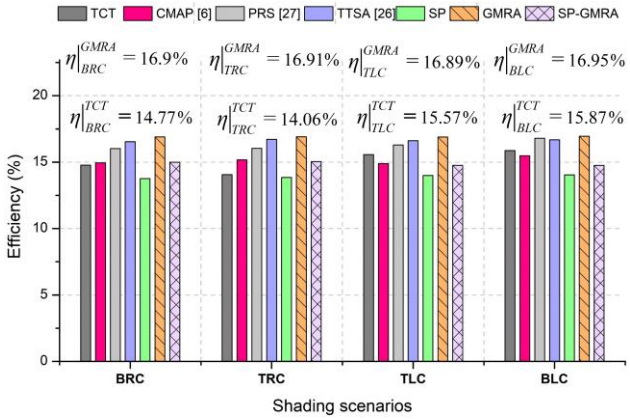


Fig. 9. Comparative efficiency yield of various PV configurations for a 9×9 PV array under different PSCs.

reconfiguration methods. The efficiency values, $\eta|_x^{TCT}$ and $\eta|_x^{GMRA}$ corresponding to TCT and GMRA configurations respectively under different shading scenarios, 'x', are also highlighted for clarity. The proposed GMRA yields the highest efficiency under all considered shading scenarios as evident from Table III. On average, 12.28 % higher efficiency is attained under all considered shading conditions using the proposed GMRA compared to the conventional TCT arrangement. SP-GMRA results in 7% higher efficiency compared to the traditional SP configuration when average values under all shading schemes were considered.

B. Techno-economic Analysis

The energy yield calculations for TCT, CMAP [6], PRS [27], TTSA [26], and the proposed GMRA technique are performed based on total power extracted after reconfiguration, considering full effective sun hours. The evaluation is performed on a 9×9 PV array of capacity 4.05 kW considering 4 different aforementioned shading scenarios (refer Table I for module specifications). Each shading scenario is assumed to occur for a duration of 10 minutes consecutively within an effective sun hour period of 8 hours, as depicted in Table IV. Furthermore, the revenue generated by the plant is estimated by considering a tariff rate of Rs. 15 per unit. As evident from Table V, the proposed reconfiguration scheme leads to the highest energy yield and therefore generates the highest annual

TABLE IV
Hourly Variation in Shading Patterns For Techno-economic Analysis

Hour	Interval					
	0-10 min	10-20 min	20-30 min	30-40 min	40-50 min	50-60 min
9 -10	BRC	TRC	TLC	BLC	BRC	TRC
10-11	TLC	BLC	BRC	TRC	TLC	BLC
11-12	BRC	TRC	TLC	BLC	BRC	TRC
12-13	TLC	BLC	BRC	TRC	TLC	BLC
13-14	BRC	TRC	TLC	BLC	BRC	TRC
14-15	TLC	BLC	BRC	TRC	TLC	BLC
15-16	BRC	TRC	TLC	BLC	BRC	TRC
16-17	TLC	BLC	BRC	TRC	TLC	BLC

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

TABLE V
ENERGY YIELD ANALYSIS FOR VARIOUS RECONFIGURATION METHODS

Hour	PV reconfiguration schemes				
	TCT	CMAP	PRS	TTSA	GMRA
9 -10	2748.5	2788.5	3007.8	3078.6	3139
10-11	2837.5	2799.5	3058.1	3089.3	3159.5
11-12	2748.5	2788.5	3007.8	3078.6	3139
12-13	2837.5	2799.5	3058.1	3089.3	3159.5
13-14	2748.5	2788.5	3007.8	3078.6	3139
14-15	2837.5	2799.5	3058.1	3089.3	3159.5
15-16	2748.5	2788.5	3007.8	3078.6	3139
16-17	2837.5	2799.5	3058.1	3089.3	3159.5
Daily energy extracted (kW-h)	22.3	22.3	24.2	24.6	**25.2**
Yearly energy extracted (kW-h)	8139.5	8139.5	8833	8979	**9198**
Total yearly income generated@ Rs 15/kW-h	122,092.5	122,092.5	132495	134685	**137970**
Cumulative yearly income generated in $	1492.0	1492.0	1619.13	1645.89	**1686.04**

revenue of $1686.04 surpassing the TCT, CMAP [6], PRS [27], and TTSA [26] methods. It is important to note that in both the TCT and the proposed GMRA arrangements, the PV modules are arranged sequentially, resulting in less wiring dependence. On the contrary, the PV arrangements such as CMAP [6] and TTSA [26], where modules can occupy any position within the array, lead to much higher cabling requirements. In the proposed GMRA arrangement, modules are repositioned within the column with fixed electrical connections, necessitating increased wiring (209×module length) for a 9×9 PV array compared to the wiring (81×module length) required for the same array size using TCT. The cost components considered for the system include the PV array with supporting structures ($ 5,167) and connecting wires ($ 1.54/m). Considering all installation costs and also factoring in a maintenance cost (2 % of the installation costs) for the PV system, the payback period for both TCT and the proposed GMRA is estimated. The techno-economic analysis revealed that the proposed system can recoup the invested funds within a time frame of about 3.2 years, which is shorter compared to the 3.56 years required with the TCT arrangement. So, the proposed reconfiguration arrangement is less costly (compared to electronic reconfigurable techniques that require a lot of sensors and switches), practically feasible, and lucrative in terms of annual energy savings for large PV plant installations.

VI. CONCLUSION

This work proposes a new generalized static SD strategy designated as GMRA and SP-GMRA, for a TCT and SP-connected PV array, respectively, for optimal power extraction under PSCs. To verify the applicability of the proposed scheme to different array sizes, initially, simulation testing is done on a large 9×9 PV array, followed by hardware validation on symmetrical 5×5 and 6×4 PV arrangements. Four benchmark performance metrics i.e. FF, ML, ER, and power enhancement were identified for the performance evaluation of the proposed scheme with popular and various recently published state-of-the-art reconfiguration methods. Simulation results revealed that the proposed GMRA and SP-GMRA achieved 12.76 % and 7.10 % higher FF than TCT and SP arrangements, respectively. Hardware implementation of GMRA revealed 29.55 % and 22.3 % enhanced average power extraction from a 5×5 and 6×4 TCT connected PV array. Additionally, the average power enhancement of 7.93 % and 10.9 % respectively is achieved for the aforementioned SP arrangements using SP-GMRA. The proposed scheme exhibits the best SD with optimal global peak power point under common shade occurrences. Simulation and experimental results show similar trends, thus validating the superiority of the proposed scheme under different PSCs.

REFERENCES

[1] J. T. Bialasiewicz, "Renewable energy systems with photovoltaic power generators: Operation and modeling," *IEEE Trans. Ind. Electron.*, vol. 55, no. 7, pp. 2752–2758, 2008.

[2] Y. Singh, B. Singh, and S. Mishra, "An Uninterruptable PV Array-Battery Based System Operating in Different Power Modes with Enhanced Power Quality," *IEEE Trans. Ind. Electron.*, vol. 69, no. 4, pp. 3631–3642, 2022.

[3] M. Thuis et al., "A Comparison of Emerging Nonfluoropolymer-Based Coextruded PV Backsheets to Industry-Benchmark Technologies," *IEEE J. Photovoltaics*, vol. 12, no. 1, pp. 88–96, 2022.

[4] I. S. Millah, R. K. Subroto, Y. W. Chang, K. L. Lian, and B. R. Ke, "Investigation of Maximum Power Point Tracking of Different Kinds of Solar Panels under Partial Shading Conditions," *IEEE Trans. Ind. Appl.*, vol. 57, no. 1, pp. 17–25, 2021.

[5] M. Kumar, "Enhanced Solar PV Power Generation under PSCs Using Shade Dispersion," *IEEE Trans. Electron Devices*, vol. 67, no. 10, pp. 4313–4320, 2020.

[6] V. M. R. Tatabhatla, A. Agarwal, and T. Kanumuri, "A Chaos Map based Reconfiguration of Solar Array to Mitigate the Effects of Partial Shading," *IEEE Trans. Energy Convers.*, vol. 37, no. 2, pp. 811–823, 2021.

[7] S. S. Reddy and C. Yammani, "Odd-Even-Prime pattern for PV array to increase power output under partial shading conditions," *Energy*, vol. 213, p. 118780, 2020.

[8] H. S. Sahu and S. K. Nayak, "Extraction of Maximum Power from a PV Array under Nonuniform Irradiation Conditions," *IEEE Trans. Electron Devices*, vol. 63, no. 12, pp. 4825–4831, 2016.

[9] N. Femia, G. Petrone, G. Spagnuolo, and M. Vitelli, "Optimization of perturb and observe maximum power point tracking method," *IEEE Trans. Power Electron.*, vol. 20, no. 4, pp. 963–973, 2005.

[10] F. Liu, S. Duan, F. Liu, B. Liu, and Y. Kang, "A variable step size INC MPPT method for PV systems," *IEEE Trans. Ind. Electron.*, vol. 55, no. 7, pp. 2622–2628, 2008.

[11] H. Patel and V. Agarwal, "Maximum power point tracking scheme for PV systems operating under partially shaded conditions," *IEEE Trans. Ind. Electron.*, vol. 55, no. 4, pp. 1689–1698, 2008.

[12] S. Makhloufi and S. Mekhilef, "Logarithmic PSO-Based Global/Local Maximum Power Point Tracker for Partially Shaded Photovoltaic Systems," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 10, no. 1, pp. 375–386, 2022.

[13] Z. Zhao et al., "Hierarchical Pigeon-Inspired Optimization-based MPPT Method for Photovoltaic Systems under Complex Partial Shading Conditions," *IEEE Trans. Ind. Electron.*, vol. 69, no. 10, pp. 10129–10143, 2021.

[14] M. Etezadinejad, B. Asaei, S. Farhangi, and A. Anvari-Moghaddam, "An Improved and Fast MPPT Algorithm for PV Systems under Partially Shaded Conditions," *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 732–742, 2022.

[15] V. M. R. Tatabhatla, A. Agarwal, and T. Kanumuri, "Performance Improvement by Mitigating the Effects of Moving Cloud Conditions," *IEEE Trans. Power Electron.*, vol. 36, no. 4, pp. 4214–4223, 2021.

[16] L. Gao, R. A. Dougal, S. Liu, and A. P. Iotova, "Parallel-connected solar PV system to address partial and rapidly fluctuating shadow conditions," *IEEE Trans. Ind. Electron.*, vol. 56, no. 5, pp. 1548–1556, 2009.

[17] D. Nguyen and B. Lehman, "An adaptive solar photovoltaic array using model-based reconfiguration algorithm," *IEEE Trans. Ind. Electron.*, vol. 55, no. 7, pp. 2644–2654, 2008.

[18] M. Z. Shams El-Dein, M. Kazerani, and M. M. A. Salama, "Optimal photovoltaic array reconfiguration to reduce partial shading losses," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 145–153, 2013.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Power Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TPEL.2023.3307813

[19] P. Srinivasa Rao, G. Saravana Ilango, and C. Nagamani, "Maximum power from PV arrays using a fixed configuration under different shading conditions," *IEEE J. Photovoltaics*, vol. 4, no. 2, pp. 679–686, 2014.
[20] B. I. Rani, G. S. Ilango, and C. Nagamani, "Enhanced power generation from PV array under partial shading conditions by shade dispersion using Su Do Ku configuration," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 594–601, 2013.
[21] H. S. Sahu, S. K. Nayak, and S. Mishra, "Maximizing the Power Generation of a Partially Shaded PV Array," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 4, no. 2, pp. 626–637, 2016.
[22] G. Sai Krishna and T. Moger, "Optimal SuDoKu Reconfiguration Technique for Total-Cross-Tied PV Array to Increase Power Output under Non-Uniform Irradiance," *IEEE Trans. Energy Convers.*, vol. 34, no. 4, pp. 1973–1984, 2019.
[23] G. Sagar, D. Pathak, P. Gaur, and V. Jain, "A Su Do Ku puzzle-based shade dispersion for maximum power enhancement of partially shaded hybrid bridge-link-total-cross-tied PV array," *Sol. Energy*, vol. 204, no. September 2019, pp. 161–180, 2020.
[24] B. Dhanalakshmi and N. Rajasekar, "Dominance square-based array reconfiguration scheme for power loss reduction in solar PhotoVoltaic (PV) systems," *Energy Convers. Manag.*, vol. 156, no. September 2017, pp. 84–102, 2018.
[25] D. S. Pillai, J. Prasanth Ram, M. Siva Sai Nihanth, and N. Rajasekar, "A simple, sensorless and fixed reconfiguration scheme for maximum power enhancement in PV systems," *Energy Convers. Manag.*, vol. 172, no. July, pp. 402–417, 2018.
[26] V. R. Krishnan, F. Blaabjerg, A. Sangwongwanich, and R. Natarajan, "Twisted Two-Step Arrangement for Maximum Power Extraction from a Partially Shaded PV Array," *IEEE J. Photovoltaics*, vol. 12, no. 3, pp. 871–879, 2022.
[27] N. Ramachandra and R. Natarajan, "Productivity Enhancement of Total-Cross-Tied PV Array Under Static Shade Conditions," *IEEE Trans. Power Electron.*, vol. 38, no. 3, pp. 4053–4060, 2023.
[28] X. Fang, Q. Yang, S. Member, and W. Yan, "Switching Matrix Enabled Optimal Topology Reconfiguration for Maximizing Power Generation in Series–Parallel Organized Photovoltaic Systems," *IEEE Syst. J.*, vol. 16, no. 2, pp. 2765–2775, 2022.

Ranjeet Singh received the B.Tech. degree in Electrical Engineering from G.B. Pant Engineering College Ghurdauri, Pauri Garhwal, India, in 2011, the M.Tech. degree in Renewable Energy Systems from National Institute of Technology, Kurukshetra, India, in 2016. He is currently working toward the Ph.D. degree with Delhi Technological University, Delhi, India. His research interests include modeling and characterization of solar cells, diagnostic and monitoring techniques for photovoltaic devices and systems, PV reconfiguration for enhanced performance and output power yield.

Vinod Kumar Yadav (Senior Member, IEEE) received the B.Tech. degree in electrical engineering from the Institute of Engineering and Technology, Bareilly, India, in 2003, the M.Tech. degree in power system engineering from the National Institute of Technology, Jamshedpur, India, in 2005, and the Ph.D. degree in power system engineering from the Indian Institute of Technology, Roorkee, India, in 2011. His research interests include renewable energy systems, power system planning and optimization, distributed generations, and smart grid.

Madhusudan Singh (Senior Member, IEEE) was born in Ghazipur, India, in 1968. He received the B.Sc. degree in electrical engineering from the Faculty of Technology, Dayalbagh Educational Institute, Dayalbagh, Agra, India, in 1990, the M.E. degree in Control and Instrumentation from the University of Allahabad, Allahabad, India, in 1992, and the Ph.D. degree from the Faculty of Technology, University of Delhi, New Delhi, India, in 2006. His research interests include modelling and analysis of electrical machines, power quality in power systems, renewable energy systems, voltage control aspects of self-excited induction generators, power electronics, drives.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:04:48 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-16494-4

Optimal threshold selection for segmentation of Chest X-Ray images using opposition-based swarm-inspired algorithm for diagnosis of pneumonia

Tejna Khosla[1] · **Om Prakash Verma**[2]

Received: 15 August 2022 / Revised: 3 June 2023 / Accepted: 8 August 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Image Segmentation using thresholding is one of the most significant areas of image processing. However, the challenge lies in accurately and effectively segmenting medical images, which is a crucial step in many applications of medical image analysis. It necessitates the development of an effective and robust segmentation approach that can handle the complexity and diversity of medical images. To address this problem, we propose a novel image segmentation technique based on minimizing the cross-entropy function using a hybrid approach that combines the features of Opposition-Based Learning (OBL), Chameleon Swarm Algorithm (CSA), and Particle Swarm Optimization Algorithm (PSO). The opposition-based technique generates the initial population and improves convergence. Then, PSO and CSA are run in parallel on an unequal population set to improve the optimal results. The proposed approach, named the Opposition-based Chameleon Swarm Algorithm improved by Particle Swarm Algorithm (CSAPSO), is evaluated on twelve Chest X-Ray (CXR) images of patients for the detection of Pneumonia. It is further tested on a large data set related to COVID-19. We conducted extensive comparisons with other state-of-the-art methods and the Deep Learning Algorithms and used the performance indicators, namely Root Mean Square Error (RMSE), Peak Signal to Noise Ratio (PSNR) and Structure Similarity Index (SSIM), Classification Accuracy, Area Under Curve for evaluating the performance. The proposed approach is statically analyzed using the Friedman rank-sum test. Through the analysis, CSAPSO demonstrates better global optimal results compared to state-of-the-art techniques.

Keywords Chameleon swarm algorithm · Cross entropy thresholding · Medical imaging · Particle swarm optimization algorithm · Pneumonia · Segmentation

✉ Tejna Khosla
patoditejna@gmail.com

Om Prakash Verma
opverma.dce@gmail.com

[1] Computer Science Engineering Department, Delhi Technological University, Bawana, Delhi, India

[2] Electronic and Communication Engineering Department, Delhi Technological University, Bawana, Delhi, India

Published online: 26 August 2023

1 Introduction

Image Segmentation forms the basis of many real-world applications like computer vision, video applications, and medical imaging. In the context of medical imaging, the CXR of the patients that are ill with respiratory symptoms is indicative of pneumonia. Pneumonia causes an increase in lung density, which can be seen as whiteness in the lungs when examining the CXR. The degree of whiteness in CXR is directly proportional to the severity of pneumonia [17]. Effective segmentation of CXR images into regions of interest (ROI) is crucial for accurate detection. Among different image segmentation techniques like thresholding, edge-detection, region-based, and graph-based [29, 51], the thresholding-based method is used widely due to its simple implementation [14, 58]. The thresholding-based segmentation method divides the image into two (bi-level) or multiple (multi-level) classes based on the number of thresholds, where all the elements in a region share common properties like brightness, grey level, contrast, texture, color, etc. The Threshold segmentation method uses the gray probability and gray value to divide the image into different regions. Since the results rely purely on the threshold values, it's important to choose the right optimal value in segmentation. In multi-level thresholding, as the number of thresholds increases, the complexity of the model increases exponentially. Also, there are many challenges when performing thresholding, like, finding a proper objective function, applying the technique over the medical image, appropriate evaluation criteria, etc. Therefore, to produce a good quality segmented image, there is a need to deal with the discussed challenges.

Several multi-level thresholding techniques have been proposed so far that employ different criteria functions, such as Otsu's between-class variance [37], minimum error thresholding [26], and entropy-based thresholding [35]. Researchers have explored all of these methods over the last decades. However, entropy-based thresholding has gained much popularity. Some common entropy measures used as objective functions include Kapur's entropy [20, 21], fuzzy entropy [40], cross-entropy [30], Tsallis entropy [8], Shannon entropy [41], and Masi entropy [34]. Among these, Minimum Cross-Entropy Thresholding (MCET) function is widely utilized. An optimal threshold value segments the image based on the difference between the original image and the ideal segmented image. The challenge of increased complexity still prevails when performing multi-level thresholding. To address this, researchers have incorporated metaheuristic algorithms to enhance threshold accuracy. In [46], bilevel and multi-level image segmentation were done by maximizing Otsu's between-class variance, using the firefly algorithm, Brownian distribution, levy flight, and the gaussian distribution. Color image segmentation based on multi-level MCET using the cuckoo search algorithm was used in [39], where authors compared the proposed approach with other metaheuristic-based image segmentation techniques, and the proposed approach proved highly effective. For complex image analysis, PSO was incorporated into MCET to find the optimal thresholds during multi-level image segmentation [57]. Another metaheuristic called grey wolf optimizer was applied to thresholding problems with Otsu and Kapur's entropy as its objective function [25]. After an extensive experimental process, this approach proved to be better than thresholding done using PSO and Bacterial Foraging Optimization (BFO). In [43], Differential Evolution was applied to Tsallis-fuzzy entropy-based image thresholding, showing statistical significance over other state-of-the-art methods using Shannon entropy or fuzzy entropy.

These studies highlight the advancements in utilizing metaheuristic algorithms to improve the accuracy and effectiveness of multi-level thresholding in image segmentation. However, there is always a limitation where some optimization algorithms are good in exploring the

search space and some are good in exploitation. Several hybrid metaheuristics are proposed in the related field to strike an exploration-exploitation balance, avoid premature convergence, and have shown superior performance. A hybrid gravitational search algorithm and the genetic algorithm were used in multi-level thresholding where between-class variance and entropy functions were used as fitness functions. In this approach, a novel method of evaluating the standard deviation after fitness function calculation was employed to ensure diversity [50]. Otsu thresholding was optimized using a hybrid firefly algorithm and social spider optimization, where processing time was significantly reduced, along with good results [3]. In [18], Ewees et al. gave a hybrid of Artificial Bee Colony (ABC) and Sine Cosine Algorithm (SCA), where SCA improved the local search ability of ABC to enhance its performance in image thresholding. It is imperative to note that hybrid optimization algorithms have shown better performance as compared to individual optimization algorithms.

In this paper, we propose an efficient segmentation approach based on MCET, called CSAPSO, that utilizes the two algorithms at different stages to improve the convergence of CSA, namely; (1) OBL [52] and (2) PSO. OBL calculates the solution on the opposite side of the ordinary solutions and chooses the best solution for each search agent. Thus, it assists in improving the convergence of the algorithm to reach the global optima value. Afterward, PSO, which exploits the search space well, is used in parallel with the CSA to take care of the exploitation feature and avoid local optima, since CSA tends to give false optimal thresholds and high complexity due to the local optima trap. To avoid this, PSO is incorporated along with CSA to ensure an optimal balance between the exploration and the exploitation of the resultant hybrid algorithm. The implementation of the proposed methodology consists of three phases: (1) Initialization and pre-processing phase, (2) updating phase, and (3) the segmentation phase. In the initialization and pre-processing phase, OBL generates a random population of agents in the opposite direction. The cross-entropy function is then minimized to select the best solutions for the next phase. In the updating phase, the population is divided into unequal sub-populations and assigned to CSA and PSO in parallel to bring population diversity, overcome the potential local optima trap, and find the global optima value. The global optima value becomes the optimal threshold value. Finally, in the segmentation phase, the image is segmented using the optimal thresholds. The proposed CSAPSO algorithm demonstrates excellent performance in image thresholding, particularly in segmenting complex medical images. It outperforms CSA, PSO, and other classical and hybrid-based segmentation approaches. The performance evaluation includes comparisons using RMSE, PSNR, and SSIM metrics, and the statistical analysis confirms the effectiveness of CSAPSO. Additionally, the algorithm's applicability is tested on a large dataset related to COVID-19, and comparisons are made with popular Data Learning Algorithms like U-Net [47], SegNet [6], and DeepLab [16]. Overall, the CSAPSO algorithm provides an efficient and competitive image segmentation method, specifically for medical image analysis, with promising results for pneumonia detection and potential applications in other domains. The main contribution of this paper are as follows:

1. A new method of multi-level image thresholding has been proposed based on MCET by evaluating the optimal threshold values based on opposition-based learning, CSA, and PSO.
2. The local and global search operators of both CSA and PSO is applied in parallel to the subpopulations, after refining the search space using OBL.
3. The performance has been tested on popular gray-scale images and real-time chest x-ray images for the detection of pneumonia. Comparison is done with other state-of-the-art nature-inspired algorithms.

4. To test the robustness and generalizability, the algorithm is further tested on COVID-19 images and compared with other Deep Learning Algorithms on the basis of classification accuracy, area under curve etc.

The remaining paper is divided as follows. Section 2 gives the existing work in multi-level thresholding and optimization algorithms in the field of medical imaging. Section 3 gives the problem definition and the preliminaries of CSA and PSO. Section 4 demonstrates the proposed CSAPSO method based on multi-level cross-entropy thresholding. Section 5 gives the experimental results and analysis. Section 6 concludes the work.

2 Related work

Researchers in the medical domain have used multi-level thresholding with nature-inspired algorithms in diverse ways. In a study [32], researchers proposed a novel method of updating the context vector during the cooperation process in a dynamic way where they used Otsu as an objective function. The testing was done on the CT scans of the stomach and was found effective. However, the computational complexity of the method may limit its scalability to large datasets or real-time applications. In 2012, three-level thresholding of human head CT images was done using improved Biogeography-Based Optimization (BBO) technique, where the fuzzy entropy was the objective measure to be maximized [15]. The comparison was with original BBO, PSO, and other state-of-the-art methods. The improved BBO proved superior and more flexible than its original algorithm and other methods. However, a careful selection of parameters is required. Abbas et al. [1] proposed a refined segmentation method based on pixel-probability density function and edge detection. The technique was applied to High-Resolution Computed Tomography (HRCT) of lungs, and compared with other segmentation methods. Brain Magnetic Resonance (MR) has been a targeted area for many researchers. However, it lacked generalization and validation on other datasets. In [38], the author proposed multi-level thresholding of brain MR images by computing the 2-d histogram-based grey gradient to preserve the edges of the brain, using Shallow Swarm Optimization (SSO). The comparisons were done with other soft computing approaches. The complexity of the algorithm was high, although it was effective in performance and gave good results. But, the proposed method is not compared to the state-of-the-art segmentation techniques to demonstrate its superiority. Also, in [28], Brain MR image segmentation was done using an enhanced Shuffled Frog Leaping Algorithm. It gave good results and was qualitatively and quantitatively better than other competitive approaches but the sample size used was very small so effectiveness cannot be claimed. Bat algorithm was used in tsallis entropy-based thresholding for image segmentation [44]. It used the region growing segmentation method and extracted the texture features to final abnormalities in MR images of the brain and breast. Li et al. used partitioned and cooperative quantum-behaved PSO to optimize the Otsu thresholding parameters and applied them to optimize complex benchmark functions and four stomach CT images [33]. The results were better in terms of inter-class variance. The approach has a high computational cost, which could make it challenging to use in real-time or in circumstances where resources are few. The dataset used was limited and the algorithm lacked analysis of the parameters. Wang et al. [55] used Otsu as the objective function and tried optimizing the thresholds using improved Flower Pollination Algorithm with a random location operator. The performance was evaluated using gray-scale medical images and gave robust and effective output as compared to other algorithms.

Firefly Algorithm (FA) was used to segment lung CT images [53] and brain MRI [45]. The paper proposes emphysema classification in lung tissue CT images using Support Vector Machine (SVM) optimized by the FA. The proposed approach's computational time and complexity are not fully analysed in the research, which is crucial for real-world applications. The grey wolf optimizer was used in an artificial neural network for the classification of the MRI images dataset [4]. The hybrid approach of using gray wolf optimizer and artificial neural network improves the accuracy of classification for magnetic resonance brain images. But, the paper does not discuss the limitations of the proposed algorithm. Moreover, Soham et al. [48] gave the concept of multi-level thresholding using a multi-objective evolutionary algorithm where two entropies namely, minimum cross-entropy and Renyi entropy, were used as the objective function. It is an effective technique to segment brain tumors. In this regard, automatic segmentation of MR images was done using Crow Search Algorithm [36]. This method used minimum cross entropy as its objective function. The performance was evaluated on general as well as medical images. Ultrasound image segmentation was carried out using a multi-level Otsu thresholding method optimized using Differential Evolution Algorithm [49]. The proposed differential search algorithm-based method achieves higher segmentation accuracy compared to other methods, as demonstrated by experiments on real ultrasound images.

For the analysis and classification of diseases in healthcare systems, machine learning and data mining approaches have become increasingly popular in recent years. In [10], authors give a review of various machine learning methods used in health care, the challenges and the possible solutions. In a manner similar, [9] investigated sub-features for classification in data mining, which was helpful for enhancing the precision of disease classification models. Analytic methods for e-Health data have also been deployed using soft computing and machine learning approaches. [12] suggested a framework for applying these methods to healthcare data analysis and patient outcomes improvement. A machine learning and medical stuff model was also used by [11] to analyse classification-based projected disease, which could be useful for diagnosing and forecasting diseases in patients. However, the sample size used was small.

In this research paper, we aimed to explore the potential of nature-inspired algorithms and their hybrids for pneumonia-related diseases. We noticed a significant gap in the literature in this area, and we identified the need for a dataset of sufficient size to address this issue. Additionally, we found that previous studies lacked proper descriptions of the parameters used for tuning and measures of accuracy. To bridge this gap, we applied nature-inspired algorithms to the image segmentation of CXR images of patients for the early detection of pneumonia. Our work represents a crucial step forward in the early detection and treatment of pneumonia patients.

3 Preliminaries

3.1 Minimum cross-entropy

The formulation of multi-level thresholding using minimum cross-entropy is presented here. Minimum cross-entropy was proposed by Kullback et al. [27]. Consider two probability distributions, $P = p_1, p_2, ..., p_N$ and $Q = q_1, q_2, ..., q_N$. The cross-entropy between the

two distributions is given as (1).

$$D(P, Q) = \sum_{k=1}^{N} p_k log(\frac{p_k}{q_k}) \tag{1}$$

where $D(P, Q)$ is the information theoretic distance between the two distributions P and Q.

Li et al. [31] have implemented minimum cross-entropy in image segmentation for bilevel thresholding. Later, it extended to multi-level thresholding. Let the original image, I with gray levels L and histogram h. For multi-level thresholding with m number of thresholds namely th_1, th_2, th_3,... th_m , the cross-entropy is minimized between the original and segmented image. $m + 1$ classes are formed, namely, $C_0, C_1, C_2, ..., C_m$, where the intensity values range between $[0, th_1]$, $[th_1 + 1, th_2]$, .., and $[th_m + 1, L - 1]$ respectively. The multi-level segmented image is denoted as I' and defined as (2).

$$I' = \begin{cases} \mu_0, I(x, y) \leq th_1 \\ \mu_1, th_1 \leq I(x, y) \leq th_2 \\ . \\ . \\ \mu_m, I(x, y) \geq th_m \end{cases} \tag{2}$$

where μ_0, μ_1, μ_2,.., μ_m is the average mean of the classes $C_0, C_1, C_2, ..., C_m$, and is given as (3).

$$\mu_0 = \frac{\sum_{i=1}^{th_1} ih(i)}{\sum_{i=1}^{th_1} h(i)}, \mu_1 = \frac{\sum_{i=th_1+1}^{th_2} ih(i)}{\sum_{i=th_1+1}^{th_2} h(i)}, ...\mu_m = \frac{\sum_{i=th_m+1}^{L} ih(i)}{\sum_{i=th_m+1}^{L} h(i)} \tag{3}$$

For multi-level thresholding, the cross-entropy function is calculated as per (4).

$$D(th_1, th_2, ..., th_m) = \sum_{i=1}^{L} ih(i)log(i) - \sum_{i=1}^{th_1} ih(i)log(\mu_0) - \sum_{i=th_1+1}^{th_2} ih(i)log(\mu_1) - ... - \sum_{i=th_m+1}^{L} ih(i)log(\mu_m) \tag{4}$$

The first term of the equation is generally constant for each image; hence the function can be modified as per (5).

$$D(th_1, th_2, ..., th_m) = -\sum_{i=1}^{th_1} ih(i)log(\mu_0) - \sum_{i=th_1+1}^{th_2} ih(i)log(\mu_1) - ... - \sum_{i=th_m+1}^{L} ih(i)log(\mu_m) \tag{5}$$

The time complexity of the multi-level thresholding increases exponentially with the increase in the number of thresholds. Since the time complexity for bilevel thresholding is $O(L)$, the time complexity for multi-level thresholding would be $O(L^m)$.

3.2 Chameleon swarm algorithm

CSA is the nature-inspired algorithm developed by Braik et al [13]. It is inspired by the hunting behavior of chameleons in the forests to sustain themselves. The chameleons have the capability of rotating their eyes 360 degrees, and work independent of each other. The hunting mechanism of the chameleon consists of three distinct steps, namely; (i) tracking the

prey, (ii) pursuing the prey with their eyes, and (iii) attacking the prey with the use of the tongue. The chameleons can roam around the entire search space in search of prey and use their globular eyes for a panoramic view of their surroundings. They use their long sticky tongue to catch the prey by the phenomenon of wet adhesion and entanglement. Their suction cup-like tongue traps the prey.

The initialization of chameleons (number and dimension) in the search space is done using (6).

$$A^i = l_j + r * (u_j - l_j) \tag{6}$$

where A^i represents the initial vector of the i^{th} chameleon, l_j, and u_j are the lower and upper bounds of the space along the j^{th} dimension respectively, r is a uniformly generated random number in the range [0, 1]. The position updation strategy can be mathematically modeled using (7).

$$A_{t+1}^{i,j} = \begin{cases} A_t^{i,j} + p_1(BP_t^{i,j} - GP_t^j)r_2 + p_2(GP_t^j - A_t^{i,j})r_1 & r_i \geq Pp \\ A_t^{i,j} + \mu((u_j - l_j)r_3 + l_j^d)sgn(rand - 0.5) & r_i < Pp \end{cases} \tag{7}$$

where $A_{t+1}^{i,j}$ is the new position of i^{th} chameleon in j^{th} dimension for iteration number $t+1$, $A_t^{i,j}$ is the position in the t^{th} iteration, $BP_t^{i,j}$ is the best position of i^{th} chameleon so far, GP_t^j is the global best position in the j^{th} dimension by any chameleon, p_1,p_2 are two positive numbers to control the exploration; r_1, r_2, r_3 are the random uniform numbers in the range [0,1], Pp is the probability of perceiving the prey. The balance between exploration and exploitation is achieved by adaptive parameter tuning of μ according to (8).

$$\mu = \gamma^{(\frac{-\alpha t}{T})^\beta} \tag{8}$$

where $\gamma\alpha\beta$ are the exploration-exploitation control parameters, t is the current iteration, T is the maximum iteration. (9) gives the new position update after detecting the prey using eye rotation.

$$A_{t+1}^i = Ar_t^i + \overline{A}_t^i \tag{9}$$

where A_{t+1}^i is the new position of the chameleon, Ar_t^i is the center of the current position before rotation and $\overline{A}_t^i$ is the rotating centered coordinates.

The velocity of the chameleon's tongue when it attacks the prey is as per (10).

$$v_{t+1}^{i,j} = \omega v_t^{i,j} + c_1(GP_t^j - A_t^{i,j})r_1 + c_2(BP_t^{i,j} - A_t^{i,j})r_2 \tag{10}$$

where ω is the inertia weight and defined in (11)

$$\omega = (1 - \frac{t}{T}^{(\rho\sqrt{\frac{t}{T}})}) \tag{11}$$

where ρ is a positive no. to manage exploitation, $v_t^{i,j}$ is the current velocity of the chameleon along j^{th} dimension.

Equation (12) shows the position of the chameleon's tongue when it attacks the prey.

$$A_{t+1}^{i,j} = A_t^{i,j} + \frac{(v_t^{i,j})^2 - (v_{t-1}^{i,j})^2}{2a} \tag{12}$$

where $v_{t-1}^{i,j}$ being the previous velocity, a is the acceleration rate of the tongue's projection and represented as (13).

$$a = 2590 * (1 - e^{-log(t)}) \tag{13}$$

Algorithm 1 summarises the steps of CSA.

Algorithm 1 Chameleon Swarm Algorithm

1. Parameter definition: Pp: position update probability; r_1, r_2, r_3, r_i: random numbers between [0,1]; u: upper bound; l: lower bound; d: dimension; $\overline{A}_t^i$ centre of the current position of chameleon i at iteration t; Ar_t^i : rotating centred coordinates of chameleon i at iteration t; T: max iteration.
2. Initialize the controlling parameters.
3. Initialize the swarm of n chameleons, $\overline{A}_t^i$i , and Ar_t^i , velocity of chameleon's dropping tongue.
4. Calculate the position of chameleons.
5. While $t < T$
6. Define μ, ω, and a
7. For $i = 1ton$
8. For $j = 1tod$
9. If $r^i \geq Pp$ then
10. Update the position using equation
    $A_{t+1}^{i,j} = A_t^{i,j} + p_1(BP_t^{i,j} - GP_t^j)r_2 + p_2(GP_t^j - A_t^{i,j})r_1$
11. Else
12. Update the position using equation
    $A_t^{i,j} + \mu((u_j - l_j)r_3 + l_j^d)sgn(rand - 0.5)$
13. End if
14. End for
15. End for
16. For $i = 1ton$
17. $A_{t+1}^i = Ar_t^i + \overline{A}_t^i$
18. End for
19. For i=1 to n
20. For j=1 to d
21. Update velocity using (10)
22. Update position using (12)
23. End for
24. End for
25. Use u and l to limit the position of chameleons
26. Evaluate and update the new position of chameleons
27. $t = t + 1$
28. end while

3.3 Particle swarm optimization

PSO algorithm [22] is inspired by the social behaviour of bird flocking, and fish schooling. It is a stochastic, non-deterministic technique, and easy to perform as well. To perform the optimization, the candidate solutions are improved repeatedly over each iteration. First, the candidate solutions (particles) are randomly selected, then they are allowed to move in search space based on the equations of position and velocity. Two important parameters that are greatly affected are $Pbest$, i.e. the local best known position, and $Gbest$, i.e. the global best position. Both these variables guide the particle to move towards the global solution. The particles should converge to their local best and global best positions with every iteration.

The particle tries to update its current velocity and position, with each iteration according to (14) and (15), respectively.

$$v_i^{t+1} = wv_i^t + c_1 * r_1 * (Pbest_i - pos_i^t) + c_2 * r_2 * (Gbest - pos_i^t) \tag{14}$$

$$pos_i^{t+1} = pos_i^t + v_i^{t+1} \tag{15}$$

where, pos_i^t and v_i^t are the current position and the velocity of the i^{th} particle at iteration t, respectively. w is the weighting function (inertia), r_1 and r_2 are random coefficients between [0,1], c_1 and c_2 are the acceleration coefficients and the two learning parameters. $Pbest_i$ is the best previous position of the i^{th} particle, and $Gbest$ is the global best solution so far of all particles.

In (14), the first part i.e., wv_i^t is responsible in maintaining the current direction, and also gives the ability of exploration, the second part i.e., $c_1 * r_1 * (Pbest_i - pos_i^t)$ gives the private thinking (cognitive development) and the third part i.e., $c_2 * r_2 * (Gbest - pos_i^t)$ gives the collaborative approach (social component). The set of sequence is repeated until the stopping criteria. In this way, PSO is able to achieve both exploration and exploitation. PSO has lesser complexity as compared to other algorithms and when it comes to performance, PSO is quite efficient. It is known for good exploitation ability; however, it lacks good exploration skills. It has been used along with many algorithms for improving their exploitation abilities. The steps of PSO algorithm implementation are summarised in Algorithm 2.

Algorithm 2 Particle Swarm Optimization Algorithm

1. Initialize the controlling parameters
2. Do
3. For each particle
4. Calculate the objective of the particle
5. Update $Pbest$ whenever required
6. Update $Gbest$ whenever required
7. End for
8. Update the inertia weight
9. For each particle
10. Update the velocity by using (14)
11. Update the position by using (15)
12. End for
13. While the end condition is not fulfilled
14. Return the best global optima Gbest

4 The proposed improved chameleon swarm algorithm

4.1 Opposition based learning mechanism

OBL was proposed by Tizhoosh in 2005 [52]. The estimates and their counter estimates form the basis of this model. Let $x \in [l, u]$ be a real number. Equation (16) formulates the opposite number $\breve{x}$.

$$\breve{x} = l + u - x \tag{16}$$

This technique effectively improves the performance of the metaheuristic algorithms by enhancing population diversity. It deals with evaluating the current solutions as well as the opposition solutions along with their fitness functions also. The better one participates in the next generation. This approach applies not only to the initial solution but to the subsequent generations of search agents as well. This is extremely helpful in cases when the current solution tends to go away from the optimal solution, especially when it is in the opposite direction.

4.2 The proposed CSAPSO algorithm

Due to stochastic behavior, there is always a scope for improvement in nature-inspired algorithms. For the same reason, CSA sometimes fails to give optimum thresholds during the image segmentation process. To utilize its advantages to the fullest and avoid the limitations, we have improvised CSA in this paper by using two algorithms, namely the Opposition-based algorithm and the Particle swarm optimization algorithm, and therefore it is named CSAPSO. Our proposed method has three phases, namely; (1) Initialization and pre-processing phase, (2) updating phase, and (3) segmentation phase. These phases are described below in more detail.

4.2.1 Initialization and pre-processing phase

Initialization is an important step in an optimization algorithm since it determines the convergence of the algorithm and the final results. OBL strategy has proved its superiority in improving the convergence and quality of the solutions. The proposed CSAPSO algorithm begins by creating a population of threshold values X in dimensions n and N solutions. (17) gives the opposite solutions for the values.

$$\breve{X} = l + u - X \tag{17}$$

where u is the upper bound, and l is the lower bound of the search space. The minimum of X and $\breve{X}$ is selected for each search agent to find the best values among the population. This is shown in (18).

$$X_i = minimum(X_i, \breve{X}_i) \qquad \forall i = 1, 2, ..., N \tag{18}$$

After the best threshold values are selected for the initial population, the objective function i.e. minimum cross-entropy function is applied to them to evaluate the performance of these thresholds, using (19).

$$F(X_i) = -\sum_{i=1}^{X_1} ih(i)log(\mu_0) - \sum_{i=X_1+1}^{X_2} ih(i)log(\mu_1) - ... - \sum_{i=X_m+1}^{X_N} ih(i)log(\mu_m) \tag{19}$$

The search agent with the minimum cross-entropy value is considered the best agent. So, OBL boosts the convergence mechanism and achieves the global optima faster.

4.2.2 Updating phase

In this phase, two subpopulations are created from the population by applying the modulus function on the iteration number. If the current iteration is odd, the population is divided into $\frac{1}{3}^{rd}$ and $\frac{2}{3}^{rd}$ subpopulations. If the current iteration is even, the population is divided into $\frac{2}{3}^{rd}$ and $\frac{1}{3}^{rd}$ subpopulations. In either case, the first subpopulation is assigned to CSA and the second sub-population is assigned to PSO respectively and passed through the fitness function, i.e., the minimum cross-entropy function. The other fitness functions that are possible for image threhsolding are Otsu, Fuzzy entropy, Ranyi entropy, wavelet entropy, and Kapur's entropy function. However, in this paper, we use the minimum cross-entropy function since this function uses the mean values which remains steady and limited. So, it becomes easy to analyse the results and draw conclusions out of them. Now, the respective CSA and PSO

algorithms update the position and velocity of all search agents. For the first subpopulation, i.e., the one with CSA, the position and velocity are updated using (20) and (21) respectively.

$$X_{t+1}^{i} = Xr_t^i + \overline{X}_t^i \tag{20}$$

$$V_{t+1}^{i,j} = \begin{cases} V_t^{i,j} + p_1(local_best_t^{i,j} - Global_best_t^j)r_2 + p_2(Global_best_t^j - V_t^{i,j})r_1 & r_i \geq Pp \\ V_t^{i,j} + \mu((u_j - l_j)r_3 + l_j^d)sgn(rand - 0.5) & r_i < Pp \end{cases} \tag{21}$$

where $Global_best$ and $local_best$ are the best global and local positions of the search agents respectively. Also, for the second subpopulation, i.e. the one with PSO, the velocity and the position are updated using (22) and (23) respectively.

$$V_i^{t+1} = wV_i^t + c_1 * r_1 * (local_best_i - X_i^t) + c_2 * r_2 * (Global_best - X_i^t) \tag{22}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{23}$$

We obtain a single set of solutions by merging two solutions. The process repeats till the termination condition. In this way, the solutions received using CSA and PSO are diverse in nature and can explore and exploit well. So, we can utilize their powerful search abilities by giving them a major and minor part of the population in an alternate way. This ensures population diversity since the population is updated in every run differently. By assigning an unequal population, we also prevent the candidate solutions from having almost the same values after every run. PSO ensures the local optima trap avoidance, and CSA retains high computational speed. In unison, both the algorithms work on improving the segmentation results.

4.2.3 Segmentation phase

After specified iterations, or until the termination condition, the final position is considered the global optima solution for the given problem. This global optima solution sets out the threshold values used in image segmentation. The solutions of the algorithm are the final threshold values. The original image uses these threshold values to form multiple segments based on the levels of thresholds. The segmentation of the image leads to dividing the image into multiple sections based on similar intensity values. The efficiency of segmentation is decided by computing the PSNR, SSIM, etc. The run-time depends on the number of iterations and the population size used for finding the optimal thresholds. the number of thresholds impacts the segmentation results and play less role in deciding the run-time of the proposed algorithm.

4.3 Implementation of the proposed CSAPSO on segmenting Chest X-Ray images for detection of pneumonia

The image segmentation using thresholding gives the classification of the image into different segments based on their similarity among them. The number of segments varies with different levels of thresholding. For distinct observations, we take a higher number of thresholds. This method proves effective for the chest x-ray images since the main criteria are to separate the affected areas of the chest from the normal ones. It could be helpful in diagnosing pneumonia at an early stage and prevent the further growth of the disease. The steps for the

implementation of the proposed algorithm for segmentation of CXR images for the detection of pneumonia is shown in Algorithm 3.

Algorithm 3 Steps for CXR image segmentation using CSAPSO

1. Input the CXR image.
2. Compute the histogram.
3. Set the initial values of the parameters of the algorithms CSA and PSO.
4. Construct the population and its opposite population using OBL.
5. Select the best population based on the values.
6. Apply the objective function and calculate the fitness value.
7. Perform CSA and PSO on unequally divided subpopulations.
8. Perform position and velocity update on each search agent and merge the final solution set.
9. When termination condition is met, the final position is the best global optimal threshold value.
10. Segment CXR image based on the optimal thresholds.

The framework of the proposed CSAPSO algorithm for image thresholding using the concept of opposition-based learning, CSA and PSO is shown in Fig. 1.

5 Experimental results and discussion

In this section, the performance of the proposed CSAPSO algorithm was tested on twelve chest x-ray images. The images were taken from [23, 24], curated by Paul Mooney and is available on Kaggle. This dataset has 5,863 images in total, 4,387 of which are training images, 585 of which are validation images, and 1,091 of which are test images. The photos are in grayscale and have a 1,024 x 1,024 pixel resolution. The dataset is separated into the "normal" and "pneumonia" classes. In contrast to the "pneumonia" class, which contains X-ray images of patients with pneumonia, the "normal" class has images of healthy patients. These Chest x-ray images along with their histogram are shown in Fig. 2. The figure shows different characteristics of the images based on their histograms.

The comparison was done with ten algorithms consisting of some popular classical as well as hybrid metaheuristics, namely; CSA, PSO, Hybrid Particle Swarm Optimization Firefly Algorithm (HFPSO) [42], Salp Swarm Algorithm (SSA) [56], Genetic Algorithm (GA) [19], Whale Optimization Algorithm (WOA) [2], Grasshopper Optimization Algorithm (GOA) [7], Gravitational Search Algorithm-Genetic Algorithm (GSA-GA) [50], Salp Swarm Optimization-Moth Flame Optimization (SSAMFO) [5], Particle Swarm Optimization-Artificial Bee Colony-Ant Colony Optimization algorithm (PSO+ABC+ACO) [54]. All the test images used for experimental evaluation were simulated thirty times using each multilevel thresholding algorithm. We experimented with the level=5, 10, 15, and 20, and the objective function to be minimized was the cross-entropy function. The results are analyzed visually as well as statistically. Visual results give the qualitative analysis and statistical results give the quantitative analysis. The experiments were done using Matlab.

The parameter settings were the same for all the algorithms as mentioned in their respective papers. The number of candidate solutions considered in each iteration is based on the population size. The search space can be better explored with a larger population, but computational cost may go up. New candidate solutions are produced from the current population using the mutation rate and crossover rate in CSA. The crossover rate regulates the likelihood of joining two solutions to produce a new one, whereas the mutation rate regulates the

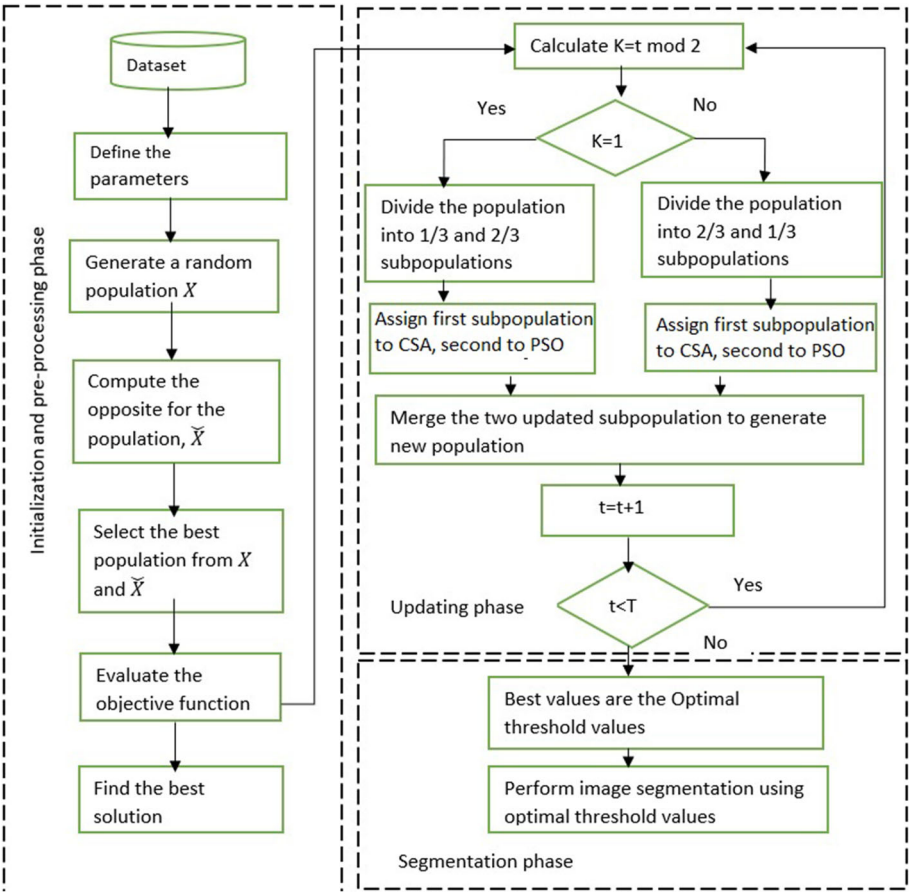
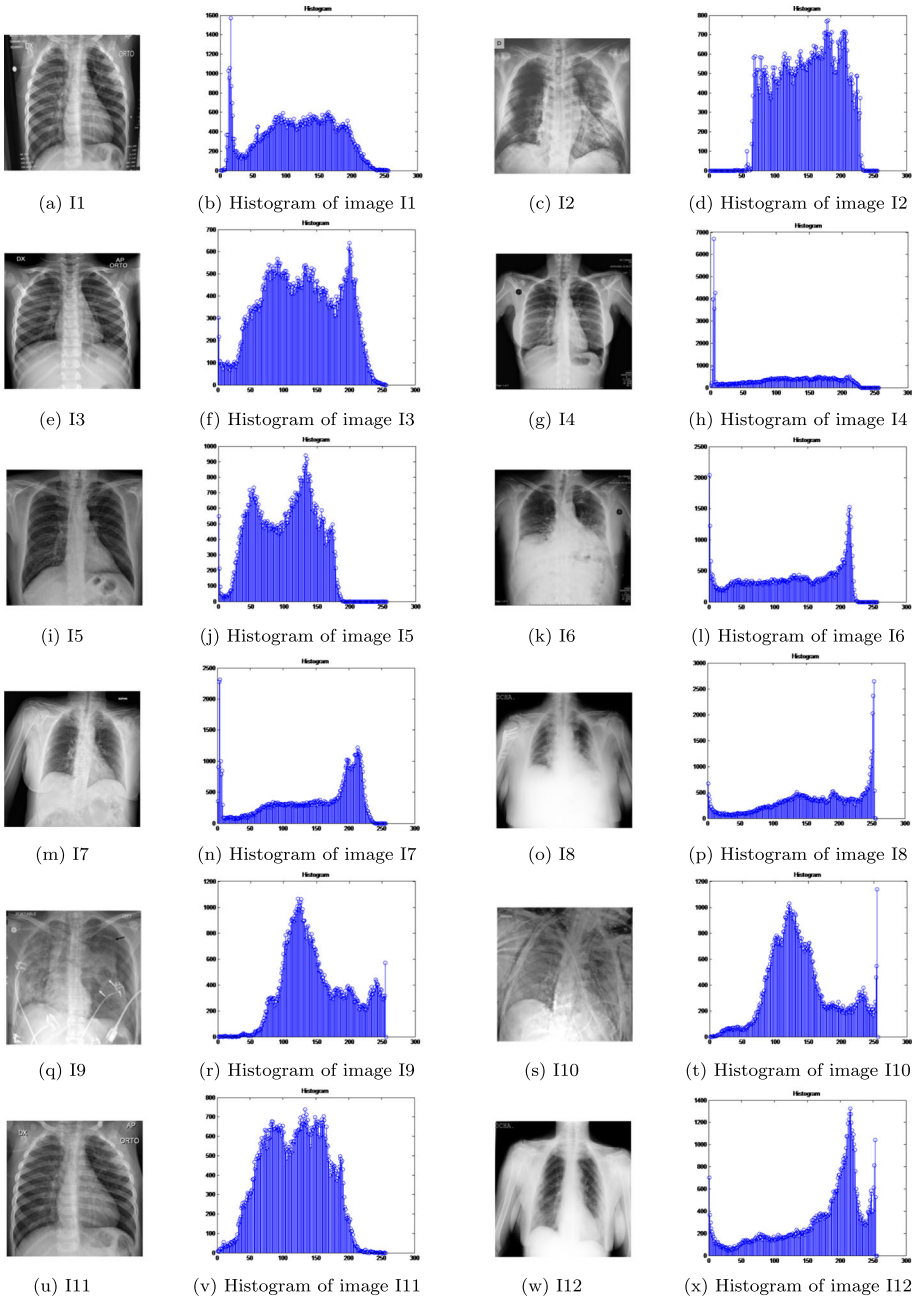


Fig. 1 Framework of the proposed CSAPSO algorithm for image thresholding using the concept of opposition-based learning, CSA and PSO

likelihood of each parameter in a solution being randomly modified. The balance between the particle's previous velocity and the social and cognitive factors in the current iteration is controlled by the inertia weight. The attraction of the particle towards its personal best solution and the best solution discovered by the swarm is controlled, respectively, by the particle's cognitive and social components. Depending on how difficult the problem is that needs to be addressed, different values for these parameters may be considered ideal. In both CSA and PSO, parameter tuning is frequently carried out by a trial-and-error process in which various parameter settings are evaluated on the problem at hand and the best performing one is chosen. The selection of parameters impact how well these algorithms perform in contrast to other optimization algorithms like the Salp Swarm Algorithm, Genetic Algorithm, Wolf Optimisation Algorithm, Grasshopper Optimisation Algorithm, etc. Finding the ideal collection of parameters is essential for getting good results because each algorithm has its own set of variables that can be tuned. To choose the optimal algorithm for a particular optimisation problem, it is important to carefully set the parameters of each method and compare their performances.For transparency and better result comparison, the two main



(a) I1 (b) Histogram of image I1 (c) I2 (d) Histogram of image I2

(e) I3 (f) Histogram of image I3 (g) I4 (h) Histogram of image I4

(i) I5 (j) Histogram of image I5 (k) I6 (l) Histogram of image I6

(m) I7 (n) Histogram of image I7 (o) I8 (p) Histogram of image I8

(q) I9 (r) Histogram of image I9 (s) I10 (t) Histogram of image I10

(u) I11 (v) Histogram of image I11 (w) I12 (x) Histogram of image I12

Fig. 2 Original Chest X-Ray Images and their Histograms

parameters, i.e., population size and maximum iteration are set to 30 and 500, respectively for all the algorithms. The other parameter values for CSAPSO are mentioned as follows: - $Pp = 0.1, \gamma = 1.0, \alpha = 3.5, \beta = 3.0, p_1 = 0.25, p_2 = 1.50, \rho = 1$

The main aim of this paper is quality segmentation results and consistency. There are several parameters used to test the quality of segmentation. We have used the following parameters as performance indicators:

1. Root Mean Square Error (RMSE) works on the intensity values and gives the strength of the final constructed image, defined as per (24).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_r}(I_{i,j} - Is_{i,j})^2}{N_r * N_c}} \tag{24}$$

   where I and Is refers to the original and segmented images of size $N_r * N_c$.
2. Peak Signal-to-Noise ratio (PSNR) gives the accuracy of the final segmented image, defined as per (25).

$$PSNR = 20 * log_{10}(\frac{255}{RMSE}) \tag{25}$$

3. Structural Similarity Index (SSIM) gives the resemblance between original image and the thresholded image. It denotes the structural information degradation and is defined as per (26).

$$SSIM(I, Is) = \frac{(2\mu_I \mu Is + c_1)(2\sigma I, Is + c_2)}{(\mu_I^2 + \mu_{Is}^2 + c_1)(\sigma_I^2 + \sigma_{Is}^2 + c_2)} \tag{26}$$

   where μ_I and μ_{Is} are the mean intensities of I and Is, respectively, and σ_I and σ_{Is} are the standard deviations of the original image I and segmented image Is, respectively. The co-variance between I and Is is denoted by $\sigma_(I, Is)$. Constants c_1 and c_2 are set to 6.5025 and 58.52252, respectively [2].

Apart from these performance indicators, Friedman rank-sum test at 5% significance level was performed to see the difference between the proposed algorithm and other algorithms.

Figure 3 shows the segmentation results on CXR images at threshold levels 5,10,15, and 20. The proposed CSAPSO algorithm successfully segmented the chest x-rays at various threshold levels. The results at threshold levels 10, 15, and 20 are better at threshold level 5. At each threshold level, the degree of whiteness in the x-rays is compared to normal chest x-rays to detect the severity of Pneumonia. As discussed, Pneumonia is identified by the degree of whiteness in the CXR. The proposed CSAPSO algorithm gave appropriate thresholds for segmentation as perceived visually from the figure.

Figure 4 shows the segmented images for image I11 by CSAPSO and other algorithms used for comparison at threshold level 20. From the comparative analysis, it is observed that all the algorithms successfully segment the images. However, for the same threshold level (in this case, 20), CSAPSO gives better image quality as compared to other metaheuristics. The dark and light regions of the chest images are clearer. The thoracic area of the chest is accurately perceived, and the identification of affected areas in x-ray images becomes distinct. Hence, from the visual analysis, CSAPSO proves to be a better algorithm as compared to its counterparts and other state-of-art algorithms for the effective segmentation of the images.

To exhibit Pneumonia, two CXR images I5 and I11 were taken at threshold level 10 segmented using CSAPSO as shown in Fig. 5. Both I5 and I11 are the CXR of persons affected by Pneumonia. The grey-colored section, denoted by the red arrow, is the affected area, whereas the black-colored section, denoted by the green arrow, is the unaffected area. The segmented image created using the thresholding technique displays the infected portions and the severity. From the figure, it is seen that there are multiple lesions or gray portions in image I11 that show how severe the disease is. The image I5 has lesser lesions due to the blacker portion and less grey portion, which shows the beginning of Pneumonia. Hence

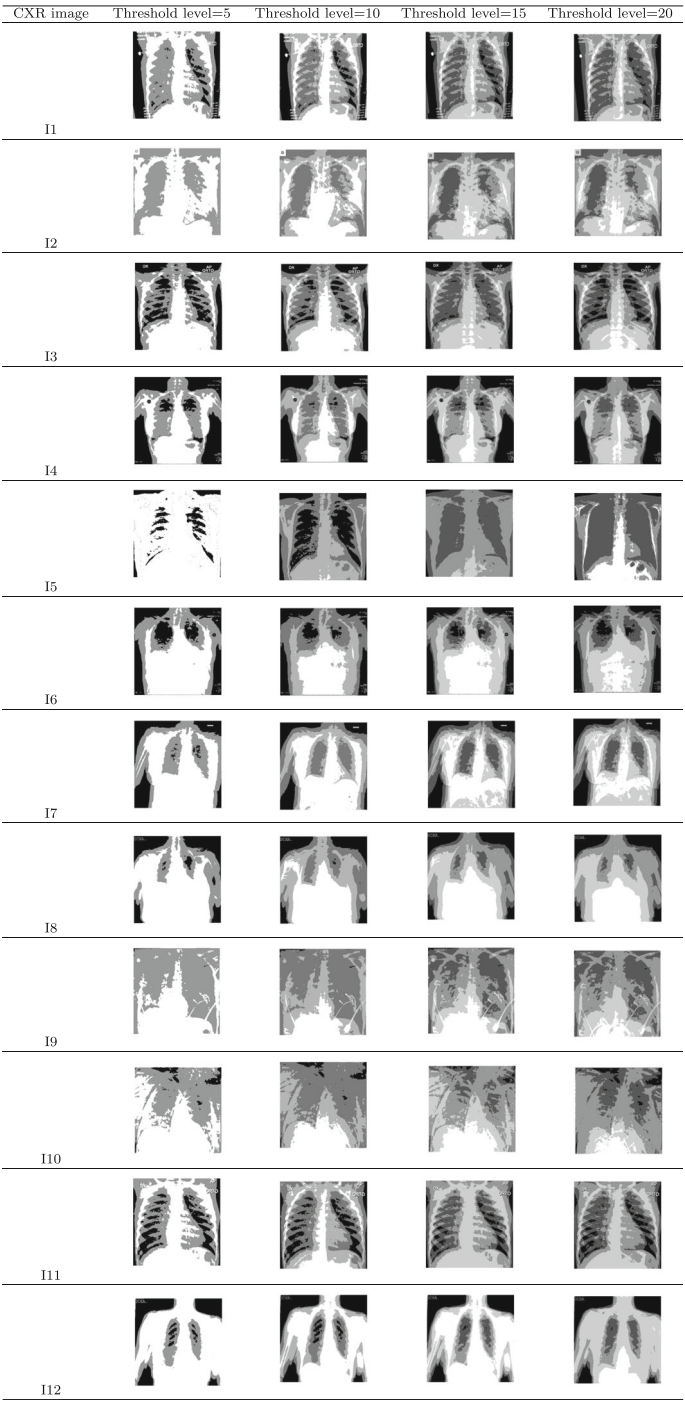
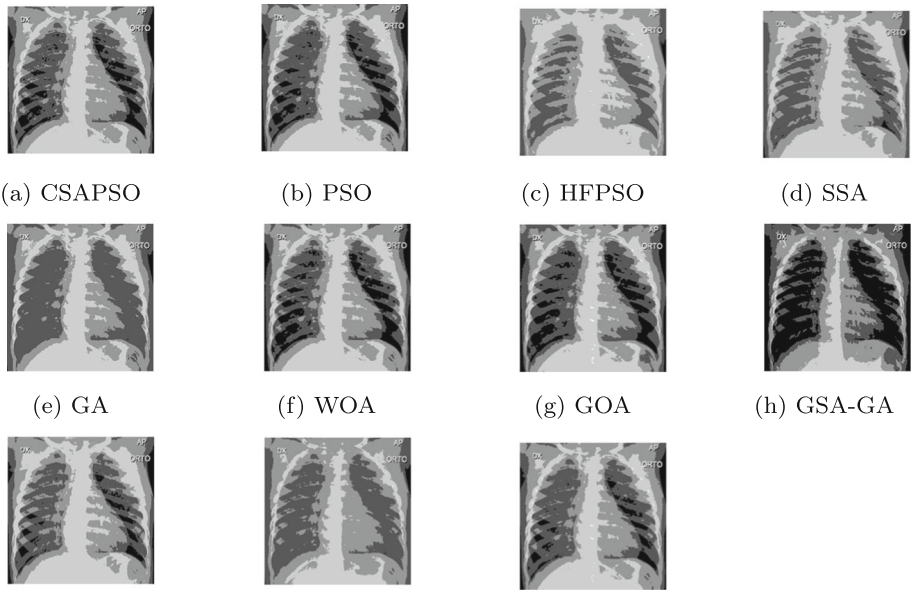


Fig. 3 The segmented CXR images (I1-I12) using CSAPSO for threshold level 5,10,15, and 20



(a) CSAPSO (b) PSO (c) HFPSO (d) SSA

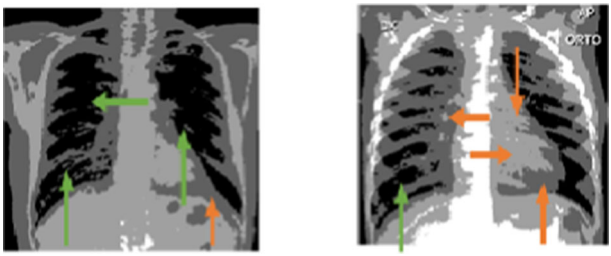
(e) GA (f) WOA (g) GOA (h) GSA-GA

(i) SSAMSO (j) PSO+ABC+ACO (k) CSA

Fig. 4 Segmented CXR images obtained by each algorithm for image I11 at threshold level 20

by segmenting the images using optimal threshold values, we can quickly know the impact of the disease on the lungs. However, the consideration is that the threshold levels chosen should be such that the affected and unaffected areas are distinctly visible. Hence, depending on the type of problem, we choose an appropriate threshold level.

To analyze the efficiency and accuracy of the proposed algorithm, fitness values, PSNR, SSIM, and CPU time are considered. Table 1 shows the fitness function values obtained using the proposed CSAPSO algorithm and other state-of-the-art algorithms. The best results are marked in bold. The fitness results give the quality of the threshold values and the accuracy of the results. It is observed from the table that CSAPSO stood first in delivering the best fitness values for most of the cases. The ranking was followed by SSAMFO and HFPSO. For some cases of lower threshold levels, SSAMFO gave better results than CSAPSO. The quality of the algorithm at higher dimensions is accessed using the higher threshold values.



(a) I5 (b) I11

Fig. 5 Illustration of pneumonia affected and unaffected areas in CXR images I5 and I11 segmented using CSAPSO

Table 1 Fitness values obtained using each algorithm for the CXR images

Test images	Levels	CSAPSO	PSO	HFPSO	SSA	GA	WOA	GOA	GSA-GA	SSAMFO	PSO+ABC+ACO	CSA
I1	5	16.84	17.79	17.65	17.67	17.72	17.08	17.66	17.41	17.83	17.69	17.28
	10	19.99	20.65	20.81	20.66	20.28	20.32	20.14	20.07	20.46	20.87	20.15
	15	28.01	20.98	28.75	28.05	28.55	28.63	28.37	28.54	27.77	28.52	28.06
	20	31.88	31.12	30.97	31.96	31.24	31.62	31.06	31.65	31.27	31.29	31.83
I2	5	17.99	18.65	18.57	18.58	18.12	18.16	18.98	18.51	18.28	18.34	18.45
	10	22.03	22.78	21.96	22.09	22.91	22.73	22.59	22.18	22.06	22.13	22.52
	15	26.87	27.26	27.65	27.71	27.56	27.77	27.68	27.48	27.15	27.82	27.59
	20	29.94	30.25	30.64	30.92	30.48	30.83	30.79	30.45	30.47	30.34	30.86
I3	5	15.83	15.23	15.69	15.16	15.77	15.31	15.41	15.37	14.86	15.02	15.12
	10	18.95	19.27	19.89	19.12	19.21	19.24	19.63	19.91	19.36	19.72	19.28
	15	26.95	27.75	27.63	27.04	27.73	27.94	27.28	27.61	27.22	27.15	27.53
	20	31.88	32.04	32.87	32.45	32.27	32.41	32.11	32.56	32.14	32.19	32.12
I4	5	10.19	10.48	9.87	10.76	10.32	10.27	10.74	10.22	10.18	10.59	10.86
	10	15.97	16.74	16.35	16.83	16.22	16.25	16.79	16. 32	16.45	16.56	16.28
	15	28.97	29.53	29.11	29.15	29.34	29.78	29.67	29.52	29.68	29.22	29.18
	20	35.61	35.32	34.95	35.77	35.04	35.11	35.89	35.01	35.64	35.19	35.26
I5	5	16.43	15.91	16.83	16.12	16.72	16.88	16.39	16.09	16.11	16.84	16.49
	10	23.97	24.51	24.69	24.45	24.35	24.27	24.26	24.86	24.92	24.23	24.62
	15	29.81	30.31	30.44	30.67	30.11	30.25	30.67	30.47	30.17	30.29	30.56
	20	39.33	39.24	39.26	39.74	39.45	39.38	39.85	39.61	38.92	39.88	39.25
I6	5	14.92	15.13	15.22	15.72	15.84	15.46	15.63	15.85	15.55	15.17	15.51
	10	19.96	20.46	20.16	20.41	20.07	20.27	20.39	20.89	20.35	20.08	20.63
	15	28.49	27.99	28.37	28.34	28.15	28.24	28.76	28.85	28.31	28.29	28.67
	20	32.31	32.14	32.29	32.13	32.66	32.29	32.86	32.47	31.95	32.36	32.21

Table 1 continued

Test images	Levels	CSAPSO	PSO	HFPSO	SSA	GA	WOA	GOA	GSA-GA	SSAMFO	PSO+ABC+ACO	CSA
I7	5	17.98	18.11	18.33	18.52	18.94	18.83	18.77	18.57	18.09	18.29	18.39
	10	22.11	22.17	22.86	22.67	22.47	22.06	22.73	22.55	21.89	22.79	22.83
	15	27.95	28.46	28.24	28.74	22.16	22.29	22.94	22.55	22.68	22.96	22.71
	20	33.72	34.47	34.66	34.59	34.64	34.23	34.13	34.53	34.14	34.18	34.28
I8	5	18.85	19.08	19.57	19.11	19.73	19.33	19.36	19.68	19.06	19.29	19.29
	10	26.03	27.33	27.45	27.87	27.59	27.06	27.29	27.67	27.95	27.24	27.86
	15	36.92	37.23	37.19	37.51	37.25	37.72	37.28	37.67	37.01	37.24	37.87
	20	40.08	40.05	40.39	40.76	40.15	40.17	40.26	40.85	39.97	40.83	40.24
I9	5	16.89	17.35	17.38	17.53	17.05	17.24	17.48	17.23	17.61	17.13	17.41
	10	18.88	19.23	19.12	19.35	19.04	19.11	19.78	19.45	19.09	19.26	19.79
	15	21.95	22.79	22.57	22.33	22.14	22.57	22.49	22.12	22.18	22.45	22.69
	20	28.45	28.18	27.97	28.23	28.55	28.38	28.79	28.23	28.15	28.91	28.16
I10	5	20.58	20.01	20.55	20.29	20.11	19.94	20.42	20.23	20.31	20.17	20.38
	10	26.97	27.03	27.65	27.49	27.07	27.28	27.13	27.41	27.27	27.64	27.86
	15	34.95	35.08	35.56	35.13	35.38	35.15	35.23	35.11	35.29	35.74	35.44
	20	39.99	40.58	40.67	40.78	40.16	40.94	40.08	40.83	40.17	40.14	40.76
I11	5	15.96	16.84	16.05	16.58	16.12	16.17	16.44	16.23	16.52	16.63	16.31
	10	24.18	24.61	24.64	23.87	24.06	24.76	24.42	24.28	24.15	24.23	24.45
	15	28.88	29.59	29.22	29.07	29.16	29.16	29.01	29.22	29.38	29.27	29.55
	20	31.99	32.27	32.19	32.56	32.85	32.33	32.05	32.74	32.69	32.56	32.79
I12	5	18.93	19.36	19.27	19.55	19.68	19.21	19.17	19.78	19.03	19.88	19.75
	10	22.19	22.49	22.64	22.22	22.57	22.27	22.08	22.47	21.96	22.11	22.84
	15	28.98	29.97	29.33	29.81	29.64	29.32	29.35	29.17	29.48	29.43	29.82
	20	40.95	41.02	41.22	41.95	41.83	41.92	41.24	41.19	41.38	41.66	41.62

Table 2 Results of the PSNR values obtained using each algorithm for the CXR images

Test images	Levels	CSAPSO	PSO	HFPSO	SSA	GA	WOA	GOA	GSA-GA	SSAMFO	PSO+ABC+ACO	CSA
I1	5	14.0628	10.8354	13.9534	11.7352	12.2131	12.7482	13.1128	10.1836	12.7384	11.2848	11.1242
	10	19.8747	18.8356	16.2821	15.6273	15.4823	17.3293	17.8495	18.3342	17.2434	18.1243	17.1233
	15	21.6753	20.1729	20.1243	19.3826	18.3863	17.3864	19.1284	18.7239	18.2173	17.1294	19.3284
	20	22.4947	20.2361	21.2924	21.1287	20.2193	19.1282	21.1283	22.1923	20.2123	19.9875	19.3947
I2	5	6.0044	5.2328	8.2322	7.1294	4.2354	10.3217	3.2734	6.2353	9.3728	8.3758	8.3748
	10	19.4746	18.8273	18.2383	16.1739	15.0363	20.2381	24.2739	22.1836	19.9256	21.0982	22.1293
	15	21.931	20.9237	23.2378	20.9273	19.7389	18.8271	19.7637	18.7382	20.1294	21.1947	18.095
	20	20.5011	24.8492	22.6943	22.2848	21.1859	21.5826	19.9356	20.4537	19.8363	18.8537	19.7517
I3	5	8.0064	7.9204	7.8263	5.1523	6.284	6.1893	7.1729	6.284	7.2383	7.293	5.2398
	10	19.9724	11.2838	19.0009	18.3789	17.8365	12.7496	15.2398	13.2739	15.284	18.9475	17.8446
	15	22.1094	21.1813	21.3819	20.8361	19.9371	21.2817	22.0961	21.2819	20.1297	19.9172	19.6601
	20	22.7307	21.0128	20.1299	21.8123	20.1281	18.1289	17.1297	18.1289	19.128	20.1298	11.0372
I4	5	8.0182	10.9372	10.6384	10.1128	9.1738	5.1389	11.979	11.1238	10.1829	11.1289	5.5568
	10	22.0849	21.281	21.0119	20.0893	15.2173	17.4463	19.0263	18.9354	16.4583	18.3749	20.9371
	15	21.6611	21.0368	20.9364	20.1284	19.2179	18.1279	19.1278	20.1289	21.2379	20.1283	16.1289
	20	21.7592	20.1289	19.2178	18.1289	20.1289	20.9464	19.1289	20.1299	20.9623	20.9172	20.6152
I5	5	13.4419	10.9217	8.1289	9.8261	9.8127	12.8124	13.8217	14.1281	14.9836	11.8736	9.7362
	10	13.8327	15.9238	15.1289	12.1289	10.9217	12.9837	16.0939	14.3782	15.9462	15.8329	14.2367
	15	19.1451	18.8462	17.8372	17.3658	16.2178	16.3278	18.9438	18.4367	17.4832	17.9384	16.3742
	20	17.8841	16.7328	15.8543	15.7384	10.74328	13.3438	12.7388	11.9584	10.0043	12.8291	13.9321
I6	5	10.3729	10.8329	11.0942	10.8452	9.6839	9.2654	11.4773	10.7491	9.3376	8.2653	10.6382
	10	17.2738	18.9472	15.2378	15.9462	17.3682	17.2356	19.339	15.7294	16.9372	18.7378	19.0328
	15	22.4181	21.8923	22.0128	20.9832	19.7823	17.7237	17.0449	19.7823	20.0129	19.2389	18.3374
	20	21.374	20.0219	19.2189	19.4378	18.8329	17.7348	18.7438	19.4839	19.1193	18.7328	19.3278

Table 2 continued

Test images	Levels	CSAPSO	PSO	HFPSO	SSA	GA	WOA	GOA	GSA-GA	SSAMFO	PSO+ABC+ACO	CSA
I7	5	10.4958	11.7382	11.3829	10.8329	9.3728	10.3829	10.0992	11.8491	12.7329	9.5849	11.5748
	10	17.1655	16.9548	16.6237	15.8943	16.8473	17.0473	17.1431	16.8437	15.2479	16.6237	15.7238
	15	23.0606	21.3879	20.0573	20.8337	20.8462	19.3278	18.3782	18.0684	21.7492	22.8362	22.0582
	20	21.9988	20.0438	20.0582	19.9974	19.7294	18.3728	17.3759	16.34649	16.4826	19.3729	18.2749
I8	5	9.8493	8.3682	9.4378	9.2327	6.3482	8.2382	10.738	8.4389	8.3278	8.4389	9.9562
	10	16.3492	15.4832	13.9347	15.9582	15.8352	16.2857	15.2834	15.9274	15.9248	15.9284	16.10934
	15	19.6393	18.3283	18.3278	16.0678	18.0473	17.4926	19.5493	18.9326	19.1837	18.4337	17.3278
	20	18.4268	19.2857	18.3281	17.3659	18.9472	18.4652	17.4822	18.4825	19.5135	18.9362	19.1736
I9	5	7.2372	5.8731	6.0288	6.8803	5.9438	6.3278	7.5934	7.1107	7.3728	6.2301	7.1901
	10	16.1167	15.3728	15.8392	15.7382	16.0483	15.4378	15.9663	14.3735	15.9261	14.2368	15.9272
	15	19.8327	18.3628	18.3725	19.4382	19.1378	18.1389	18.9363	19.7462	18.37682	17.8366	18.3725
	20	23.8867	22.8563	22.4826	22.8741	23.6382	23.8462	22.4926	22.9472	21.9461	20.4392	20.3641
I10	5	9.1063	8.6472	8.9472	8.2834	9.0173	8.1268	8.3634	8.1635	8.9371	7.2718	6.1562
	10	18.2437	17.3728	17.1832	16.4735	16.5724	15.3619	18.4707	17.37619	16.3289	15.3826	15.3715
	15	18.3795	19.3867	19.3712	18.3712	20.0163	18.8462	20.1308	17.0562	16.9462	17.8526	18.8472
	20	19.7059	18.6381	19.0431	18.3728	18.8462	18.9371	19.4381	18.6482	18.4527	18.3268	17.3692
I11	5	12.4645	10.7362	11.9562	11.2545	10.2548	9.5282	10.5241	12.1835	11.7361	10.1356	11.5731
	10	16.546	16.2644	15.3108	15.0388	14.0931	14.0162	14.2673	13.6395	15.3687	15.3216	15.8534
	15	18.7655	17.3261	17.9654	17.6435	16.6432	16.3405	15.2549	14.8452	17.7352	17.3561	14.0351
	20	22.0763	21.8239	21.9463	20.7129	19.8351	21.9456	21.8831	21.5593	20.6691	21.8274	21.4374
I12	5	9.1733	8.6271	8.3627	8.2738	7.2562	8.56281	9.6599	8.1994	7.1763	7.237	8.3287
	10	18.9142	16.6342	17.6581	16.9473	16.5098	15.7459	16.6392	16.4262	17.6372	17.9435	18.2849
	15	20.4071	19.3487	19.3748	18.3728	18.3946	17.3657	18.3265	17.3267	17.9462	18.4825	17.3711
	20	20.7885	19.8462	19.3649	18.3906	18.4628	17.0462	17.4474	18.3628	17.9846	18.8361	19.8437

Table 3 Results of the SSIM values obtained using each algorithm for the CXR images

Test images	Levels	CSAPSO	PSO	HFPSO	SSA	GA	WOA	GOA	GSA-GA	SSAMFO	PSO+ABC+ACO	CSA
I1	5	0.5058	0.4378	0.4312	0.4639	0.4915	0.3715	0.4829	0.4871	0.4123	0.3821	0.2634
	10	0.6839	0.6283	0.6812	0.6245	0.6477	0.6245	0.5892	0.6182	0.6934	0.6528	0.6539
	15	0.7058	0.7037	0.7894	0.7363	0.7152	0.7637	0.7193	0.7473	0.7542	0.7079	0.7024
	20	0.8947	0.8827	0.8884	0.8007	0.8292	0.8057	0.8615	0.8633	0.8284	0.8259	0.8813
I2	5	0.4765	0.4623	0.4387	0.4225	0.4264	0.4923	0.4885	0.4916	0.4219	0.4657	0.4685
	10	0.6913	0.5528	0.5764	0.5025	0.5764	0.5174	0.5122	0.5172	0.6182	0.5336	0.5981
	15	0.6987	0.6364	0.6139	0.6036	0.6842	0.6652	0.6837	0.6208	0.6715	0.6759	0.6843
	20	0.7893	0.7665	0.7563	0.7126	0.7943	0.7677	0.7615	0.7201	0.7943	0.7842	0.7684
I3	5	0.5242	0.5061	0.5525	0.5308	0.5521	0.5582	0.5652	0.5224	0.5934	0.5226	0.5202
	10	0.6937	0.6784	0.6761	0.6768	0.6166	0.6777	0.6311	0.6876	0.6743	0.6527	0.6587
	15	0.7372	0.7803	0.7149	0.7505	0.7002	0.7226	0.7016	0.7544	0.7951	0.7783	0.7943
	20	0.8934	0.8458	0.8325	0.8177	0.8912	0.8687	0.8162	0.8292	0.8446	0.8422	0.8401
I4	5	0.3969	0.3002	0.3266	0.3142	0.3562	0.3774	0.3205	0.3422	0.3568	0.3476	0.3499
	10	0.4329	0.4539	0.4964	0.4804	0.4068	0.4182	0.4562	0.4775	0.4852	0.4049	0.4225
	15	0.5173	0.5268	0.5153	0.5182	0.5139	0.5073	0.5155	0.5698	0.5741	0.5525	0.5237
	20	0.6316	0.6121	0.6013	0.6544	0.6899	0.6139	0.6703	0.6748	0.6987	0.6561	0.601
I5	5	0.4866	0.4722	0.4663	0.4315	0.4145	0.4813	0.4703	0.4525	0.4317	0.4322	0.4474
	10	0.5943	0.5299	0.5736	0.5609	0.5365	0.5231	0.5465	0.5273	0.5468	0.5358	0.5317
	15	0.6908	0.6906	0.6115	0.6465	0.6172	0.6302	0.6434	0.6356	0.6562	0.6848	0.6207
	20	0.7745	0.7954	0.7565	0.7956	0.7006	0.7727	0.7664	0.7163	0.7115	0.7661	0.7435
I6	5	0.5989	0.5647	0.5208	0.5432	0.5743	0.5008	0.5054	0.5471	0.5817	0.5746	0.5671
	10	0.6989	0.6082	0.6101	0.6095	0.6187	0.6706	0.6832	0.6099	0.6204	0.6318	0.6955
	15	0.7906	0.7887	0.7895	0.7852	0.7001	0.7239	0.7263	0.7271	0.7266	0.7242	0.7154
	20	0.8211	0.8213	0.8716	0.8529	0.8263	0.8402	0.8332	0.8314	0.8416	0.8391	0.8088

Table 3 continued

Test images	Levels	CSAPSO	PSO	HFPSO	SSA	GA	WOA	GOA	GSA-GA	SSAMFO	PSO+ABC+ACO	CSA
I7	5	0.4299	0.4622	0.4805	0.4687	0.4741	0.4599	0.4201	0.4463	0.4922	0.4308	0.4897
	10	0.5577	0.5803	0.5093	0.5635	0.5149	0.5856	0.5644	0.5253	0.5366	0.5009	0.5153
	15	0.5849	0.5181	0.5386	0.5437	0.5301	0.5246	0.5129	0.5532	0.5329	0.5246	0.5336
	20	0.7765	0.7201	0.7502	0.7108	0.7302	0.7298	0.7695	0.7175	0.7481	0.7533	0.7135
I8	5	0.4382	0.4442	0.4279	0.4725	0.4174	0.4131	0.4708	0.4533	0.4991	0.4205	0.4831
	10	0.5914	0.5807	0.5103	0.5719	0.5276	0.5578	0.5269	0.5325	0.5363	0.5421	0.5568
	15	0.6628	0.6297	0.6366	0.6297	0.6202	0.6814	0.6207	0.6408	0.6807	0.6998	0.6191
	20	0.7822	0.7811	0.7155	0.7454	0.7152	0.7138	0.7814	0.7602	0.7429	0.7505	0.7596
I9	5	0.4953	0.4733	0.4368	0.4205	0.4172	0.4205	0.4763	0.4724	0.4836	0.4759	0.4625
	10	0.5109	0.5772	0.5665	0.5744	0.5394	0.5139	0.5485	0.5564	0.5805	0.5496	0.5579
	15	0.6124	0.6403	0.6805	0.6957	0.6444	0.6179	0.6198	0.6613	0.6051	0.6133	0.6446
	20	0.7872	0.7277	0.7306	0.7817	0.7152	0.7375	0.7274	0.7802	0.7673	0.7715	0.7335
I10	5	0.5998	0.5366	0.5287	0.5029	0.5907	0.5001	0.5502	0.5918	0.5295	0.5572	0.5925
	10	0.6215	0.6113	0.6516	0.6839	0.6627	0.6814	0.6167	0.6756	0.6941	0.6319	0.6396
	15	0.7782	0.7811	0.7857	0.7816	0.7412	0.7679	0.7104	0.7397	0.7619	0.7841	0.7273
	20	0.8395	0.8401	0.8363	0.8618	0.8234	0.8323	0.8179	0.8328	0.8788	0.8361	0.8274
I11	5	0.3957	0.3179	0.3234	0.3244	0.3359	0.3485	0.3285	0.3829	0.3369	0.3239	0.3328
	10	0.4996	0.4371	0.4219	0.4916	0.4782	0.4286	0.4957	0.4972	0.4835	0.4301	0.4731
	15	0.5983	0.5373	0.5978	0.5083	0.5514	0.5708	0.5077	0.5798	0.5947	0.5073	0.5419
	20	0.6988	0.6782	0.6388	0.6956	0.6671	0.6238	0.6073	0.6869	0.6144	0.6445	0.6627
I12	5	0.5431	0.5447	0.5365	0.5671	0.5579	0.5398	0.5712	0.5301	0.5876	0.5663	0.5498
	10	0.6124	0.6892	0.6399	0.6983	0.6255	0.6519	0.6204	0.6463	0.6291	0.6395	0.6291
	15	0.7313	0.7505	0.7538	0.7477	0.7305	0.7531	0.7345	0.7467	0.7176	0.7498	0.7485
	20	0.8693	0.8342	0.8477	0.8191	0.8279	0.8669	0.8691	0.8528	0.8357	0.8283	0.8242

It is significant since real-world applications may contain several objects. One observation is that for higher threshold levels, the performance of CSAPSO was much better than at lower threshold levels. It shows that CSPSO proves to be a better optimizer than other algorithms for real-world applications.

Table 2 shows the results of the average PSNR results for all the images. From the detailed analysis of the performance at each threshold level, CSAPSO performs great for threshold levels 15 and 20. For threshold level 5, CSAPSO gave the best PSNR values for 4 cases, followed by GOA which gave the best values in 3 cases, followed by SSAMFO for 2 cases. For threshold level 10, CSAPSO gave the best PSNR values in 8 cases, followed by GOA for 4 cases. For threshold level 5, the performance of GOA and HFPSO was similar. However, it is outperformed by CSAPSO which gave the best values in 10 cases. Similar results were shown for threshold level 20, where the performance of GOA, SSAMFO, and PSO was the same, while CSAPSO outperformed others in 9 cases. Therefore, from the detailed analysis of the PSNR results, it can be seen that the performance of CSAPSO was the best, followed by GOA, and then SSAMFO. The other comparative algorithms like CSA, PSO, SSA, and WOA proved the worst for these cases. The impact of balanced exploration and exploitation can also be found in the quality of results. Since the local trap is avoided, therefore, the quality of results is significantly improved.

Table 3 shows the average SSIM values for the images obtained using the proposed algorithm and other comparative algorithms. SSIM tells the similarity between the original image and the corresponding segmented image, hence it is an important quality parameter to validate the segmentation. The results were analyzed in detail and it was found that for threshold level 5, CSAPSO and SSAMFO were highly competitive and gave the best results in 6 and 5 cases respectively. For level 10, CSAPSO stood first with the best values in 6 cases, followed by SSAMFO and HFPSO. For level 15, CSAPSO gave the best results in 5 cases out of 12, followed by HFPSO, and then SSAMFO. The proposed approach gave the best results for 58% of cases in threshold level 20, and the rest of the best results were divided among SSAMFO (17%), GA (8%), SSA (8%), and HFPSO (8 %). Overall, it was observed that CSAPSO is competitive with SSAMFO for lower threshold levels. However, in totality, CSAPSO stood first, followed SSAMFO, and then HFPSO. The performance of CSA is improved by using PSO and OBL.

Figure 6 shows the comparison of the average CPU taken by each algorithm for the first 5 CXR images over 30 runs and 500 iterations. It can be seen from the figure that the CPU time taken by CSAPSO was nearly 0.5 seconds, followed by PSO with 1 second execution time, SSAMFO which was nearly 1.3 seconds, and HFPSO with 1.5 seconds. SSA gave good results with nearly 3 seconds of execution. The algorithms like GA, GOA, and CSA took nearly 4 seconds of execution time. GSA-GA and WOA took around 5 seconds. The maximum CPU time was taken by PSO+ABC+ACO with nearly 6 seconds. The results of CDAPSO are achieved in lesser CPU time due to decreased complexity. Since PSO is a simple-to-implement algorithm, therefore when it is hybridized with CSA, it reduces the complexity of the algorithm and makes the function evaluations quicker. One more interesting observation was that the graph of CPU execution time was almost a straight line which shows that the CPU time did not change much with the change in threshold levels. The results are the same with all the algorithms.

We tested the significance and efficiency of the proposed algorithm by using the Friedman test. This test ranks the algorithms and studies the algorithm's robustness by using the median values. The higher the value, the better the rank. Also, the chief advantage is that it does not require the distribution of the dataset to be known. Table 4 presents the results of the Friedman rank-sum test. Our proposed approach obtained the best rank in the measures PSNR as well

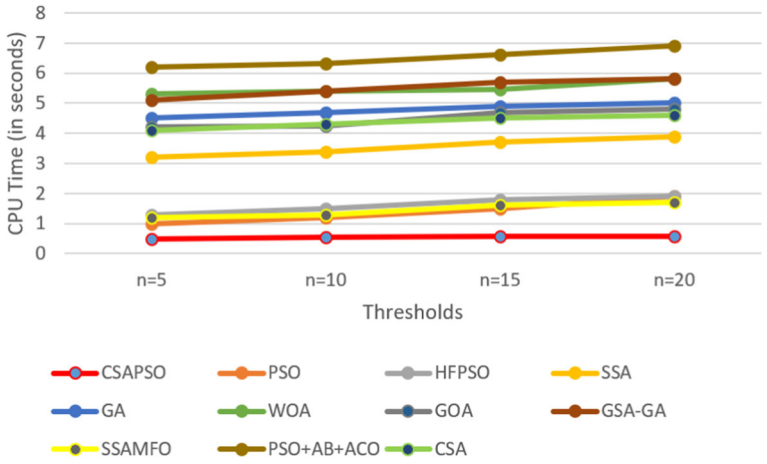


Fig. 6 Comparison of average CPU time by each algorithm for first five images

as SSIM, followed by SSAMFO and then HFPSO. The overall ranking of all the algorithms for both the measures is $CSAPSO > SSAMFO > HFPSO > PSO + ABC + ACO > PSO > CSA > SSA > GSA - GA > WOA > GOA > GA$. Therefore, CSAPSO proves to be statistically significant as compared to other algorithms for solving real-world problems.

Another series of experiment was conducted on a large data set on COVID-19.[1] In this series, the three deep learning models, namely U-Net [47], SegNet [6], and DeepLab [16] were trained over a number of parameters. Figure 7 shows the graphical evaluation of all proposed X-ray image classifiers with four metrics of precision, recall, specificity and f1-score. The positive cases of Pneumonia and COVID-19 are confirmed while using U-Net and CSAPSO. The mis-classification chances are more in case of DeepLab and SegNet. So, from the figure it can be concluded that the right classification is achieved in case of U-Net and CSAPSO. Four expected outcomes of the confusion matrix are defined as follows by comparing the true labels of tested images and predicted outcomes of deep learning: The term True Positive (TP) demonstrated the accurate identification of chest diseases. A count of healthy instances is referred to as true negatives (TN). False positive (FP) refers to an error in which a test result falsely suggests the existence of a chest disease when the disease is not present. False negative (FN) refers to an error in which a test result fails to reveal the presence of a chest disease. From the results of the confusion matrix, the following five fundamental performance measures for deep image classifiers are as follows:Accuracy: This measure evaluates the performance of the classifier. It is calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

The precision recall or sensitivity, specificity, and f1-score are calculated as follows:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = Senstivity = \frac{TP}{TP+FN}$$

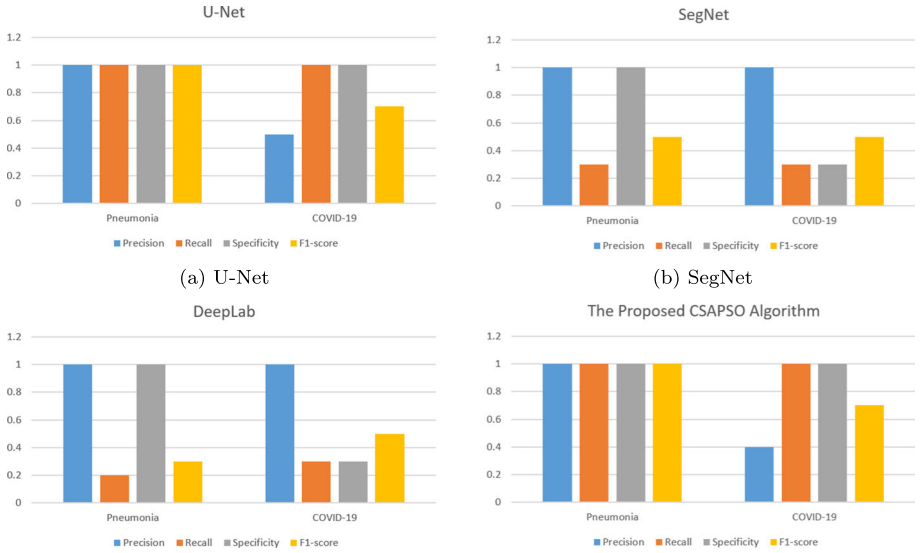
$$Specificity = \frac{TN}{TN+FP}$$

$$F1 - Score = \frac{2(PrecisionXRecall)}{Precision+Recall}$$

[1] The images were taken from https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711 and https://github.com/ieee8023/covid-chestxray-dataset

Table 4 Results of the Friedman test obtained using each algorithm for the CXR images

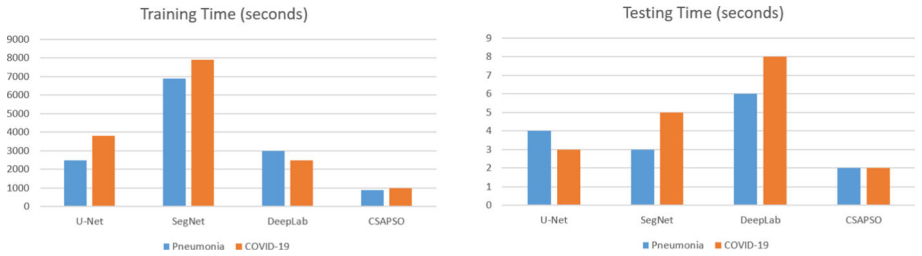
	CSAPSO	PSO	HFPSO	SSA	GA	WOA	GOA	GSA-GA	SSAMFO	PSO+ABC+ACO	CSA
PSNR	8.63	6.98	7.94	6.39	4.97	5.54	5.14	5.72	8.12	7.56	6.75
SSIM	7.28	6.19	6.97	5.93	3.42	5.13	3.95	5.52	7.01	6.68	6.05



(a) U-Net (b) SegNet

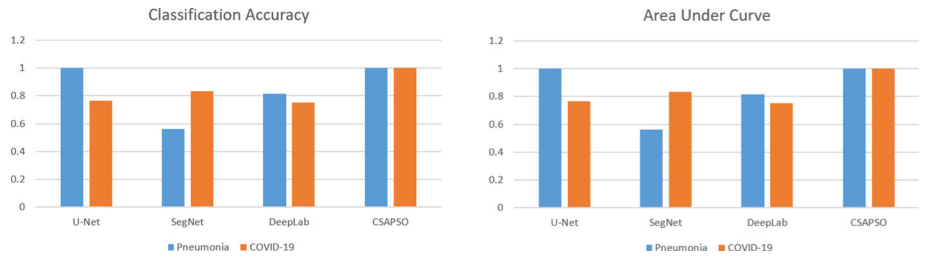
(c) DeepLab (d) CSAPSO

Fig. 7 Evaluation results of deep learning classifiers to identify positive cases of pneumonia and COVID-19 diseases



(a) Training Time (b) Testing Time

Fig. 8 Computation times including training and testing times for the Deep Learning Classifiers



(a) Accuracy (b) Area Under Curve

Fig. 9 Classification accuracy and area under curve (AUC) for the Deep Learning Classifiers

The comparative training and testing time analysis is shown in Fig. 8. The training time ranges from 900 seconds for CSAPSO to 7800 seconds for SegNet. On the other hand, the testing time ranges from 2 seconds for CSAPSO to 8 seconds for DeepLab. It shows an advantage of using CSAPSO with minimum training time of 900 seconds and testing time of 2 seconds. A higher training and testing time results in increased complexity.

Figure 9 shows the testing accuracy and area under curve for all the image classifiers. The best scores of accuracy and area under curve are achieved by CSAPSO, i.e. 0.9, followed by U-Net. SegNet and DeepLab followed the performance of CSAPSO and U-Net. The tabular comparison of different deep learning classifiers are shown in Table 5. From the table, it is seen that U-Net and CSAPSO, both classifies Pneumonia with 99.98% accuracy. However, when the dataset of COVID-19 is considered, CSAPSO outperforms U-Net with classification accuracy of 99.93% where U-Net has accuracy of 76.45%. The classifiers SegNet and DeepLab showed an average performance for both the datasets.

From the above series of experiments, we confirm the superiority of the proposed algorithm over other state-of-the-art algorithms. Our hybrid method, CSAPSO, combines the Chameleon Swarm Algorithm (CSA), Particle Swarm Optimisation (PSO), and Opposition-Based Learning (OBL), allowing for quick convergence to the global optimum. This helps in reducing the amount of compution needed for image segmentation. CSAPSO utilizes the parallel execution of CSA and PSO on different sub-populations, which enhances optimization and improves computational efficiency. This ensures population diversity, which helps in avoiding local optima trap and produces superior segmentation outcomes. Medical image analysis is only one of the many image segmentation jobs that CSAPSO can be used for. It has been evaluated for the detection of pneumonia on Chest X-Ray (CXR) pictures, and it has demonstrated promising outcomes on COVID-19-related datasets.The Friedman test results proved that CSAPSO is statistically significant as compared to other algorithms. While our proposed method, CSAPSO, combines metaheuristic algorithms rather than deep

Table 5 Comparative analysis of different deep learning classifiers

Classifier	Total Parameters(10^6)	Trainable Parameters(10^6)	Non-Trainable Parameters(10^6)	Classification Accuracy	
				Pneumonia	COVID-19
U-Net	14.93	14.93	0	99.98	76.45
SegNet	20.76	20.71	0.05	56.14	83.23
DeepLab	58.12	58.1	0.02	81.34	75.12
CSAPSO	3.54	3.51	0.03	99.98	99.93

learning techniques, however, our study also compared three deep learning techniques with our suggested method to segment medical images, and the results based on Accuracy, Precision, Recall, Specificity, and F-1 score, revealed that all four techniques were effective on the datasets we evaluated. There were some variations in performance, with CSAPSO outperforming U-Net, SegNet and Deep Lab. Our approach does, however, have a few drawbacks that need to be addressed in further research. First, the existing architecture relies on time-intensive manual annotations for training. Second, a rather small dataset was used to train and test the suggested network, which would have limited its potential to generalise to other datasets or real-world events. Finally, the suggested method may not be appropriate for other medical imaging tasks because it is created specifically for the detection and localization of pneumonia in CXR images. The findings imply that deep learning methods for medical image segmentation show considerable promise, but further study is required to choose the most appropriate method for particular imaging modalities and applications. The study also emphasises the need for larger datasets and standardised evaluation measures to enable reliable comparison of various approaches.

6 Conclusion

In this paper, we present the improvement of the CSA algorithm, named CSAPSO, through the integration of two other algorithms: PSO and OBL. The resultant algotrithm, namely CSAPSO, is a multi-level thresholding-based image-segmentation method based on minimum cross-entropy principle. CSAPSO minimizes the cross-entropy function by utilizing optimal threshold values obtained using CSA and PSO in a hybrid fashion. OBL is employed to enhance convergence towards the global optimal solution. Additionally, CSAPSO incorporates the division of the population into unequal subpopulations to ensure population diversity. The two subpopulations are run in parallel, with one using CSA algorithm and the other utilizing the PSO algorithms. This allows the algorithm to leverage the global and local features of both algorithms simultaneously.

The proposed CSAPSO algorithm is implemented for the segmentation of twelve real-time CXR images and COVID-19 X-ray images. The experimental results are compared with state-of-the-art algorithms such as CSA, PSO, HFPSO, SSA, GA, WOA, GOA, GSA-GA, SSAMFO, PSO+ABC+ACO, as well as various deep learning algorithms including U-Net, SegNet, and DeepLab. Performance indicators such as RMSE, PSNR, SSIM, accuracy, and area under the curve are used for evaluation. The statistical significance of the results is tested using the Friedman rank test at a 5% significance level.

The experimental results demonstrate that the proposed approach is highly competitive at the 5-threshold level and outperforms other algorithms at the threshold levels of 10, 15, and 20. The achieved segmentation accuracy, as indicated by the high PSNR and SSIM values, further highlights the effectiveness of the proposed method. The lower values of standard deviation indicate the stability of the algorithm, which is a crucial measure as many algorithms exhibit instability by producing different results with each iteration. The results of the Friedman rank-sum test also confirm the significance of CSAPSO in comparison to other methods. Moreover, the proposed algorithm successfully applies to real-life applications by producing optimal thresholds and generating high-quality segmented images of CXR for pneumonia detection. Therefore, it proves to be valuable in the field of medical imaging, enhancing the diagnostic performance of radiologists.

In future works, CSAPSO can be further utilized to optimize other image segmentation methods. Additionally, the development of a multi-objective version of the CSAPSO-based image thresholding technique could potentially improve its performance in segmenting CXR images.

Data Availability The datasets generated and analysed during this study are available in the following repository: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711 and https://github.com/ieee8023/covid-chestxray-dataset.

Declarations

Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Abbas Q, Khan MTA, Farooq A, Celebi ME (2013) Segmentation of lungs in hrct scan images using particle swarm optimization. Int J Innov Comput Inf Control 9(5):2155–2165
2. Abd El Aziz M, Ewees AA, Hassanien AE (2017) Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation. Expert Syst Appl 83:242–256
3. Abd El Aziz M, Ewees AA, Hassanien AE (2016) Hybrid swarms optimization based image segmentation. In: Hybrid soft computing for image segmentation, Springer, pp 1–21
4. Ahmed HM, Youssef BA, Elkorany AS, Saleeb AA, Abd El-Samie F (2018) Hybrid gray wolf optimizer-artificial neural network classification approach for magnetic resonance brain images. Appl Opt 57(7):B25–B31
5. Alwerfali HSN, Abd Elaziz M, Al-Qaness MA, Abbasi AA, Lu S, Liu F, Li L (2019) A multilevel image thresholding based on hybrid salp swarm algorithm and fuzzy entropy. IEEE Access 7:181,405–181,422
6. Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(12):2481–2495
7. Bhandari AK, Rahul K (2019) A novel local contrast fusion-based fuzzy model for color image multilevel thresholding using grasshopper optimization. Appl Soft Comput 81(105):515
8. Bhandari AK, Kumar A, Singh GK (2015) Tsallis entropy based multilevel thresholding for colored satellite image segmentation using evolutionary algorithms. Expert Syst Appl 42(22):8707–8730
9. Bhuyan HK, Ravi V (2021) Analysis of subfeature for classification in data mining. IEEE Trans Eng Manag
10. Bhuyan HK, Ravi V, Brahma B, Kamila NK (2022) Disease analysis using machine learning approaches in healthcare system. Health and Technology 12(5):987–1005
11. Bhuyan HK, Sai TA, Charan M, Chowdary KV, Brahma B (2022) Analysis of classification based predicted disease using machine learning and medical things model. 2022 Second International Conference on Advances in Electrical. Computing, Communication and Sustainable Technologies (ICAECT), IEEE, pp 1–6
12. Brahma B, Bhuyan HK (2022) Soft computing and machine learning techniques for e-health data analytics. In: Connected e-Health: Integrated IoT and Cloud Computing, Springer, pp 83–104
13. Braik MS (2021) Chameleon swarm algorithm: A bio-inspired optimizer for solving engineering design problems. Expert Syst Appl 174(114):685
14. Breve F (2019) Interactive image segmentation using label propagation through complex networks. Expert Syst Appl 123:18–33
15. Chatterjee A, Siarry P, Nakib A, Blanc R (2012) An improved biogeography based optimization approach for segmentation of human head ct-scan images employing fuzzy entropy. Eng Appl Artif Intell 25(8):1698–1709
16. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Transactions on Pattern Analysis and Machine Intelligence 40(4):834–848

17. Cleverley J, Piper J, Jones MM (2020) The role of chest radiography in confirming covid-19 pneumonia. bmj 370
18. Ewees AA, Abd Elaziz M, Al-Qaness MA, Khalil HA, Kim S (2020) Improved artificial bee colony using sine-cosine algorithm for multi-level thresholding image segmentation. IEEE Access 8:26,304–26,315
19. Hammouche K, Diaf M, Siarry P (2008) A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation. Comp Vision Image Underst (CVIU) 109(2):163–175
20. Jiang Z, Zou F, Chen D, Kang J (2021) An improved teaching-learning-based optimization for multilevel thresholding image segmentation. Arab J Sci Eng 46(9):8371–8396
21. Kapur JN, Sahoo PK, Wong AK (1985) A new method for gray-level picture thresholding using the entropy of the histogram. Computer vision, graphics, and image processing 29(3):273–285
22. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of ICNN'95-international conference on neural networks, IEEE, vol 4. pp 1942–1948
23. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F et al (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(5):1122–1131
24. Kermany D, Zhang K, Goldbaum M et al (2018) Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley data 2(2)
25. Khairuzzaman AKM, Chaudhury S (2017) Multilevel thresholding using grey wolf optimizer for image segmentation. Expert Syst Appl 86:64–76
26. Kittler J, Illingworth J (1986) Minimum error thresholding. Pattern Recogn 19(1):41–47
27. Kullback S (1968) Information theory and statistics (peter smith, gloucester, ma). Information theory and statistics
28. Ladgham A, Hamdaoui F, Sakly A, Mtibaa A (2015) Fast mr brain image segmentation based on modified shuffled frog leaping algorithm. SIViP 9(5):1113–1120
29. Lee SH, Koo HI, Cho NI (2010) Image segmentation algorithms based on the machine learning of features. Pattern Recogn Lett 31(14):2325–2336
30. Li CH, Lee C (1993) Minimum cross entropy thresholding. Pattern Recogn 26(4):617–625
31. Li C, Tam PKS (1998) An iterative algorithm for minimum cross entropy thresholding. Pattern Recogn Lett 19(8):771–776
32. Li Y, Jiao L, Shang R, Stolkin R (2015) Dynamic-context cooperative quantum-behaved particle swarm optimization based on multilevel thresholding applied to medical image segmentation. Inf Sci 294:408–422
33. Li Y, Bai X, Jiao L, Xue Y (2017) Partitioned-cooperative quantum-behaved particle swarm optimization based on multilevel thresholding applied to medical image segmentation. Appl Soft Comput 56:345–356
34. Nie F, Zhang P, Li J, Ding D (2017) A novel generalized entropy and its application in image thresholding. Signal Process 134:23–34
35. Nie F, Zhang P, Li J, Ding D (2017) A novel generalized entropy and its application in image thresholding. Signal Process 134:23–34
36. Oliva D, Hinojosa S, Cuevas E, Pajares G, Avalos O, Gálvez J (2017) Cross entropy based thresholding for magnetic resonance brain images using crow search algorithm. Expert Syst Appl 79:164–180
37. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66
38. Panda R, Agrawal S, Samantaray L, Abraham A (2017) An evolutionary gray gradient algorithm for multilevel thresholding of brain mr images using soft computing techniques. Appl Soft Comput 50:94–108
39. Pare S, Kumar A, Bajaj V, Singh GK (2017) An efficient method for multilevel color image thresholding using cuckoo search algorithm based on minimum cross entropy. Appl Math Comput 61:570–592
40. Pare S, Bhandari AK, Kumar A, Singh GK (2018) A new technique for multilevel color image thresholding based on modified fuzzy entropy and lévy flight firefly algorithm. Comput Electr Eng 70:476–495
41. Pun T (1981) Entropic thresholding, a new approach. Computer Graphics and Image Processing 16(3):210–239
42. Rahkar Farshi T, K Ardabili A (2021) A hybrid firefly and particle swarm optimization algorithm applied to multilevel image thresholding. Multimedia Systems 27(1):125–142
43. Raj A, Gautam G, Abdullah SNHS, Zaini AS, Mukhopadhyay S (2019) Multi-level thresholding based on differential evolution and tsallis fuzzy entropy. Image Vis Comput 91(103):792
44. Raja NSM, Fernandes S, Dey N, Satapathy SC, Rajinikanth V (2018) Contrast enhanced medical mri evaluation using tsallis entropy and region growing segmentation. Journal of Ambient Intelligence and Humanized Computing pp 1–12

45. Raja NSM, Lakshmi PV, Gunasekaran KP (2018) Firefly algorithm-assisted segmentation of brain regions using tsallis entropy and markov random field. In: Innovations in Electronics and Communication Engineering, Springer, pp 229–237
46. Rajinikanth V, Couceiro M (2015) Rgb histogram based color image segmentation using firefly algorithm. Procedia Computer Science 46:1449–1457
47. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, pp 234–241
48. Sarkar S, Das S, Chaudhuri SS (2017) Multi-level thresholding with a decomposition-based multiobjective evolutionary algorithm for segmenting natural and medical images. Appl Soft Comput 50:142–157
49. Shao D, Xu C, Xiang Y, Gui P, Zhu X, Zhang C, Yu Z (2019) Ultrasound image segmentation with multilevel threshold based on differential search algorithm. IET Image Process 13(6):998–1005
50. Sun G, Zhang A, Yao Y, Wang Z (2016) A novel hybrid algorithm of gravitational search algorithm with genetic algorithm for multi-level thresholding. Appl Soft Comput 46:703–730
51. Tang N, Zhou F, Gu Z, Zheng H, Yu Z, Zheng B (2018) Unsupervised pixel-wise classification for chaetoceros image segmentation. Neurocomputing 318:261–270
52. Tizhoosh HR (2005) Opposition-based learning: a new scheme for machine intelligence. In: International conference on computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce (CIMCA-IAWTIC'06), IEEE, vol 1. pp 695–701
53. Tuba E, Tuba M, Simian D (2017) Support vector machine optimized by firefly algorithm for emphysema classification in lung tissue ct images. University of West Bohemia, Digital Library
54. Upadhyay P, Chhabra JK (2021) Multilevel thresholding based image segmentation using new multistage hybrid optimization algorithm. Journal of Ambient Intelligence and Humanized Computing 12:1081–1098
55. Wang R, Zhou Y, Zhao C, Wu H (2015) A hybrid flower pollination algorithm based modified randomized location for multi-threshold medical image segmentation. Bio-medical materials and engineering 26(s1):S1345–S1351
56. Wang S, Jia H, Peng X (2020) Modified salp swarm algorithm based multilevel thresholding for color image segmentation. Math Biosci Eng 17(1):700–724
57. Yin PY (2007) Multilevel minimum cross entropy threshold selection based on particle swarm optimization. Appl Math Comput 184(2):503–513
58. Yue X, Zhang H (2019) Improved hybrid bat algorithm with invasive weed and its application in image segmentation. Arab J Sci Eng 44(11):9221–9234

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Trans Indian Inst Met
https://doi.org/10.1007/s12666-023-02996-0

ORIGINAL ARTICLE

Parametric Optimization of FSWed Dissimilar Composite Joints of AA7075 and AA6061 Using RSM

R. S. Mishra[1] · **Sumit Jain**[1]

Received: 17 December 2022 / Accepted: 30 May 2023
© The Indian Institute of Metals - IIM 2023

Abstract In the current work, SiC particles-based friction stir welded composite joints of AA7075 and AA6061 were developed. The effects of processing variables (rotating speed, transverse speed and volume percentage of SiC microparticles) on the mechanical characteristics (tensile strength, microhardness and %strain) of the composite joints were examined using response surface methodology. The developed models showed that the rotating speed is the most influential parameter, the volume percentage of SiCp is the second most influential parameter, and the transverse speed is the least influential parameter. The highest tensile strength of 261 MPa, %strain of 32.4% and microhardness of 187 HV were found at rotating speed of 1100 rpm, transverse speed of 40 mm/min and 10% volume percentage of SiC particles. The optimized processing variables, including rotating speed, transverse speed and volume percentage of SiC particles, were found to be 1094 rpm, 40 mm/min and 6%, respectively. The value of multi-optimized output responses, including tensile strength, %strain and microhardness, was found to be 247 MPa, 26.9% and 176 HV, respectively.

Keywords Composite joints · SiC particles · Tensile strength · Microhardness · ANOVA

1 Introduction

In almost every industrial facility, including automotive, marine, and aerospace, the production of light and relatively robust structural components has lately experienced a transition from conventional materials to lightweight materials. The excellent physical and mechanical features of aluminum alloys include their lightweight, high strength, strong heat conductivity and superior corrosive resistance. These alloys have attracted much attention for industrial applications in practically every area. Fusion welding procedures are inappropriate for joining such alloys of low melting points because of limitations such as porosity, distortion, solidification cracking and high residual stresses [1–3]. Friction stir welding (FSW), which is solid-state by nature, seems to be the most efficient method to join various lightweight materials. The FSW technique was introduced in 1991 by "TWI" (The Welding Institute) [4]. The fabrication of metal matrix composites (MMCs) employing reinforcing particles (RPs) is a creative technique for augmenting mechanical characteristics (stiffness, hardness and strength, etc.), corrosion, and wear resistance [5–7]. Bulk and surface composites were developed using RPs-based MMCs via FSP to augment tribological and mechanical properties [8–13].

Several researchers also employed FSW to produce aluminum metal matrix (AMM) composites utilizing various micro/nano-sized RPs to ameliorate the weld properties of related aluminum alloys. Recent studies have examined the viability of fabricating MMCs with different reinforcing particles, including AA2124 using SiC particles, AA7005 using Al_2O_3 particles, AA2009 using SiC particles and AA7075/AA2024 using SiC particles [14–17]. Al_2O_3 particles outperform SiC particles and a hybrid of the two in FSW of the dissimilar AA6101 and AA1350 in terms of mechanical and wear characteristics [18]. A few scholars have optimized the processing parameters of FSW using different experimental design methodologies, including response surface methodology (RSM), Taguchi approach and grey relational analysis [19–21]. However, there is a paucity of studies related to

✉ Sumit Jain
sumitjain4@gmail.com

[1] Department of Mechanical Engineering, Delhi Technological University, New Delhi, Delhi, India

Published online: 28 July 2023

optimizing the volume percentage of RPs and FSW processing variables to multi-optimize the response variables. In the current work, the desirability function using RSM was used to identify the optimal variables setting to optimize the ultimate tensile strength (UTS), %strain and microhardness of FSWed composite joints.

A wide range of structural components for the aviation, marine and automotive industries employ AA7075 and AA6061 [22–24]. The joining of these alloys with improved weld quality will boost the adaptability of structural designs and widen the scope of their applications. Thus, in the current investigation, the potential of SiC particles (SiCp) on the mechanical properties of dissimilar FSWed composite joints of AA7075 and AA6061 was investigated. The work based on the optimization of friction stir welding parameters (rotating speed and transverse speed) along with volume percentage of RPs is a novel approach that can significantly enhance the mechanical properties of the welded joints. The optimization of the volume percentage of RPs, together with rotating speed and transverse speed for considered responses of dissimilar FSWed composite joints of AA7075 and AA6061, has not previously been described, as far as the author is concerned.

2 Material and Experimental Procedure

The plates of the alloys (AA7075 and AA6061) of size 150 × 40x6 mm were utilized to develop FSWed composite joints as delineated in Fig. 1a. The grain structure micrograph and the elemental composition, as observed by EDX (energy-dispersive X-ray spectroscopy) analysis of base alloys, are cited in Fig. 2a–d. To investigate the influence of varying volume percentages of SiCp, the grooves of different sizes (width- 0.24, 0.42, and 0.60 mm, and depth- 3 mm) were milled on the continuous faces of base plates, as delineated in Fig. 1a. The SiCp of average particle size 50 µm were utilized as reinforcement material. The SEM (scanning electron microscopy) micrograph and EDX analysis of SiCp are delineated in Fig. 2e, f. A slurry was developed by mixing SiCp with ethanol to introduce the SiCp properly in the grooves [25, 26]. In the present study, the tool tilt angle of 2° was used, and AA7075 base plates positioned on the

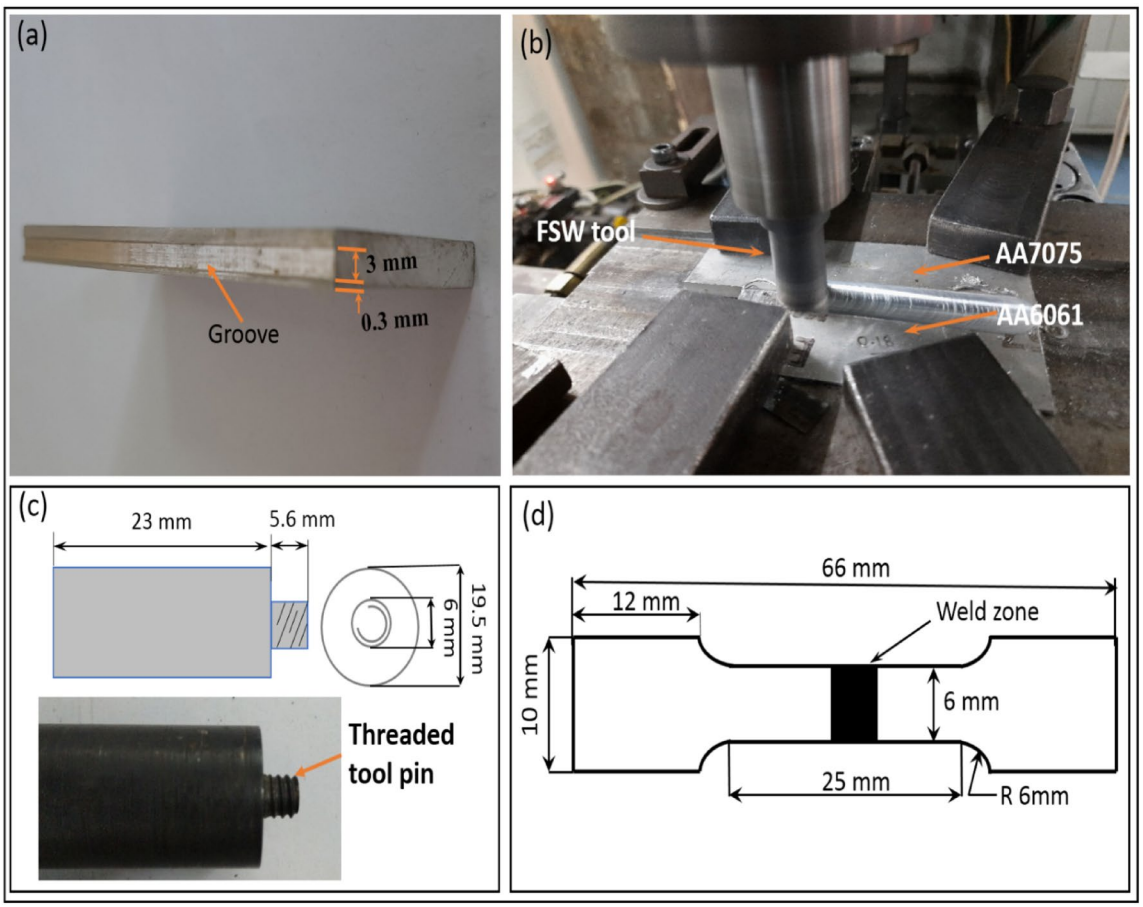


Fig. 1 **a** Base plate with groove dimensions; **b** FSW Experimental setup; **c** FSW Tool; **d** Tensile sample dimensions

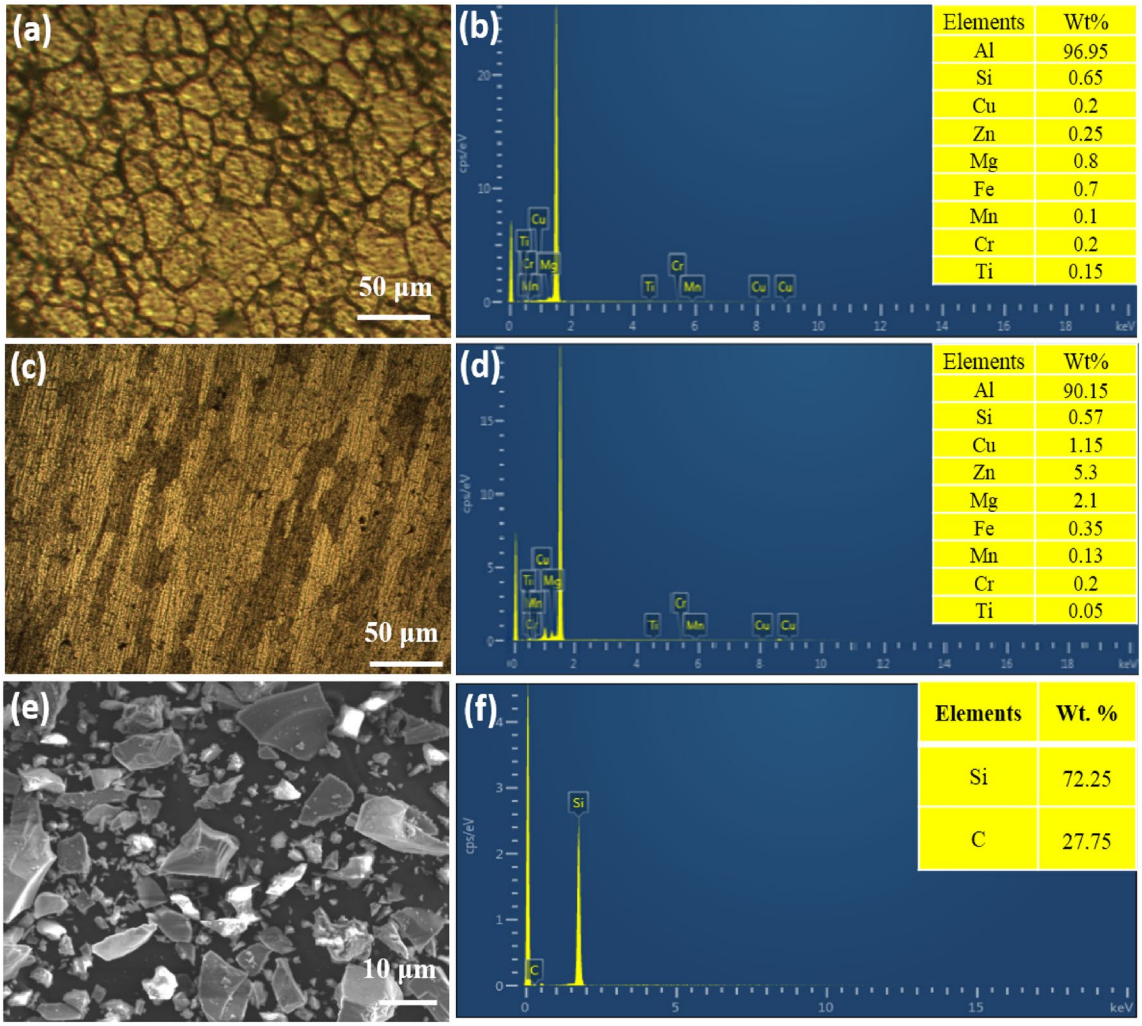


Fig. 2 **a**–**d** Grain structure and elemental composition; (**a**, **b**) AA6061, (**c**, **d**) AA7075; **e** SEM photo-image of SiCp, **f** Elemental composition of SiCp

retreating side were employed (Fig. 1b), as per the recommendations from previous studies [27, 28].

RSM is a mathematical approach for maximizing the impact of independent variables on dependent parameters. To generate the experimental design, the full factorial module of the central-composite design approach of RSM was used [29]. The experimental design of twenty trail runs was obtained with three independent processing variables: transverse speed (TS), rotating speed (RS), and volume percentages of SiCp (VP-SiCp) and three levels of each independent processing parameter. Table 1 depicts the processing parameters with their levels and considered response parameters. A threaded pin-profiled H13 steel tool was utilized to fabricate the dissimilar composite joints under different parametric settings. The dimensional characteristics of the tool utilized in this study are delineated in Fig. 1c.

The test samples for mechanical characteristics and microstructure evolution were carved out from the composite joints using wire-cut CNC EDM. The dimensions of tensile test samples (ASTM E8 standard) are delineated in Fig. 1d [30]. The average of three samples of various tests was listed. The microhardness variation in the weld region was evaluated using Vickers microhardness tester. The indents for microhardness measurement were employed at 100 gf load and 15 s dwell time. For microstructural analysis, polishing of the samples was carried out using different emery papers (200–2000 grit size) and then etched using Keller's reagent. An optical microscopy (OM) was utilized for the grain structure evolution of the stir zone, whereas a SEM was employed to examine the dispersion pattern of SiCp.

3 Results and Discussion

The grain structure and SiCp distribution pattern in the stir zone of composite joints were studied to understand the

Table 1 Mechanical characteristics of SiCp composite joints

Specimen no	Rotating speed (rpm)	Transverse speed (mm/min)	VP-SiCp (%)	UTS (MPa)	%Strain (%)	Microhardness (HV)
1	900	50	7	207	24.3	144
2	700	40	10	194	21.7	139
3	900	50	7	208	24.8	146
4	1100	50	7	244	26.7	168
5	900	40	7	219	24.3	152
6	700	60	10	205	23.2	141
7	900	50	7	208	24.5	142
8	900	50	7	203	24.2	141
9	900	50	10	221	26.8	151
10	1100	40	4	234	25.2	169
11	1100	60	10	255	29.8	181
12	900	50	7	207	24.3	145
13	900	60	7	210	25.2	147
14	900	50	7	207	24.2	144
15	1100	40	10	261	32.4	187
16	900	50	4	188	23.2	131
17	1100	60	4	219	25.1	154
18	700	60	4	180	21.8	130
19	700	40	4	185	18.8	132
20	700	50	7	195	20.2	136

consequences of RS, TS, and VP-SiCp on the mechanical and microstructural behavior. The study provides a comprehensive analysis of the results. This includes a comparison of the properties of the welded joints processed using different combinations of RS, TS and volume percentage of RPs, and the study of the underlying mechanisms that contribute to the observed changes in properties. The mean grain size of AA6061 and AA7075 (base materials) was found to be 36 µm and 47, respectively (Fig. 2a and c). The various dissimilar FSWed composite joints after FSW/SiCp are delineated in Fig. 3. In the FSWed joints, three different zones are obvious: stir zone (SZ), heat-affected-zone (HAZ) and thermo-mechanical-affected-zone (TMAZ) [31]. The grain structure of the SZ was noticed finer than that of the TMAZ and HAZ [32, 33].

3.1 Tensile Properties

The mechanical characteristics of the FSWed composite joints are mainly influenced by grain refinement, volume percentage and size of RPs, dispersion of RPs and the bonding between the AMM and RPs [34, 35]. In Table 1, the average tensile properties of all the SiCp composite joints were compiled. The base materials AA7075 and AA6061 demonstrated the UTS of 495 and 288 MPa, respectively, with corresponding %strain values of 21.6 and 18.4%.

The composite FSWed joints processed at RS of 700 rpm revealed the clusters of SiCp in the SZ and poor bonding between the SiCp and the surrounded AMM owing to inadequate material mixing. The clusters of SiCp in the SZ are obvious in Fig. 4a of specimen no- 18 processed at 700 rpm, 60 mm/min and 4% of VP-SiCp. Therefore, the mean grain size of 15.7 µm (Fig. 4b) was observed in the SZ. This indicates that the lower value of RS (700 rpm) was not providing enough strain to prevent SiCp clustering [36]. The dispersion pattern of SiCp in the SZ was found improved as RS increased [37, 38]. Uniform dispersion of SiCp and considerable material mixing was noticed due to significant strain produced at higher RS of 1100 rpm. Consequently, specimen no- 15 processed at 1100 rpm, 40 mm/min, and 10% of VP-SiCp revealed the homogeneously dispersed SiCp due to excellent material mixing caused by enhanced stirring action of the tool as depicted in Fig. 5a. The excellent bonding between SiCp and the AMM is implied by the absence of SiCp clustering. The more refined grain structure with grain size 4.6 µm (Fig. 5b) was observed in the SZ of specimen no- 15 and that only be ascribed to the presence of homogeneously distributed SiCp. The pinning effect of SiCp enhances the nucleation sites and prevents the grains' growth [39]. The presence of a high volume percentage of more homogenously dispersed SiCp led to improved grain refinement and enhanced the pinning effect. These results

Fig. 3 Welded plates of AA7075 and AA6061 embedded with SiCp

validate the influences of SiCp on the refinement of the grain structure.

The Hall–Petch equation, which states that UTS rises as grain size decreases, can explain this phenomenon [40]. The pinning effect of SiCp and DRX may be attributed to the reduced grain size of the FSWed composite joints when compared to base materials [41]. Consequently, the maximum UTS of 261 MPa was found in specimen no-15, which is nearly 90.6% of the AA6061. In contrast, the minimum UTS (180 MPa) was found in specimen no-18. The FSWed composite joints showed the higher %strain than base materials, with specimen no-15 showing a maximum %strain of 32.4%. The existence of SiCp was confirmed by the EDX analysis of the SZ of specimens no-18 and no- 15. Figures 4c and 5c depict the peaks of silicon and carbon, which validates the presence of SiCp. The EDX analysis of specimen no-18 revealed the silicon and carbon percentages of 4.08 and 2.22%, respectively, whereas the EDX analysis of specimen no- 15 revealed the silicon and carbon percentages of 8.16 and 3.11%, respectively. The varying volume percentage of SiCp was confirmed by the different percentages of silicon and carbon in specimens no-18 and no-15.

Figure 6a, b depicts the fractography of the raptured surface of specimens no- 18 and no- 15. The rapture occurred from the HAZ toward AA6061, where the microhardness was found to be minimal [42]. The shear plane with a cup-cone form was established along the edges of fractured samples [42]. The ductile failure with honeycomb dimples was found in specimen no- 18 (Fig. 6a), which provides evidence of the regularly separated attributes of cleavage and plastic deformation [43, 44]. Homogeneously dispersed SiCp and finer grain structure resulted from efficient material mixing at higher RS. The fractography of specimen no-15 exhibited the equiaxed and tiny dimples (Fig. 6b) developed by the coalescence of the micro-voids [45, 46]. The ruptured zone was identified at the HAZ, confirmed by the microhardness variance, and produced an improvement in ductility. Therefore, it was evident from examining the ruptured surfaces that adding SiCp to composite joints enhanced their tensile properties.

3.2 Microhardness

Vickers hardness tester was utilized to examine the microhardness variation in the welded area of the various FSWed composite joints, as delineated in Fig. 7a–c. The asymmetrical microhardness distribution in the welded region was induced by the irregular plastic flow in the retreating and advancing side [47]. The microhardness of the base materials AA6061 and AA7075 was found to be 96 and 170 HV, respectively. Table 1 summarizes the mean microhardness value at the midpoint of various FSWed composite joints.

The results of the microhardness tests showed that specimen no-15 exhibited a maximum microhardness of 187 HV,

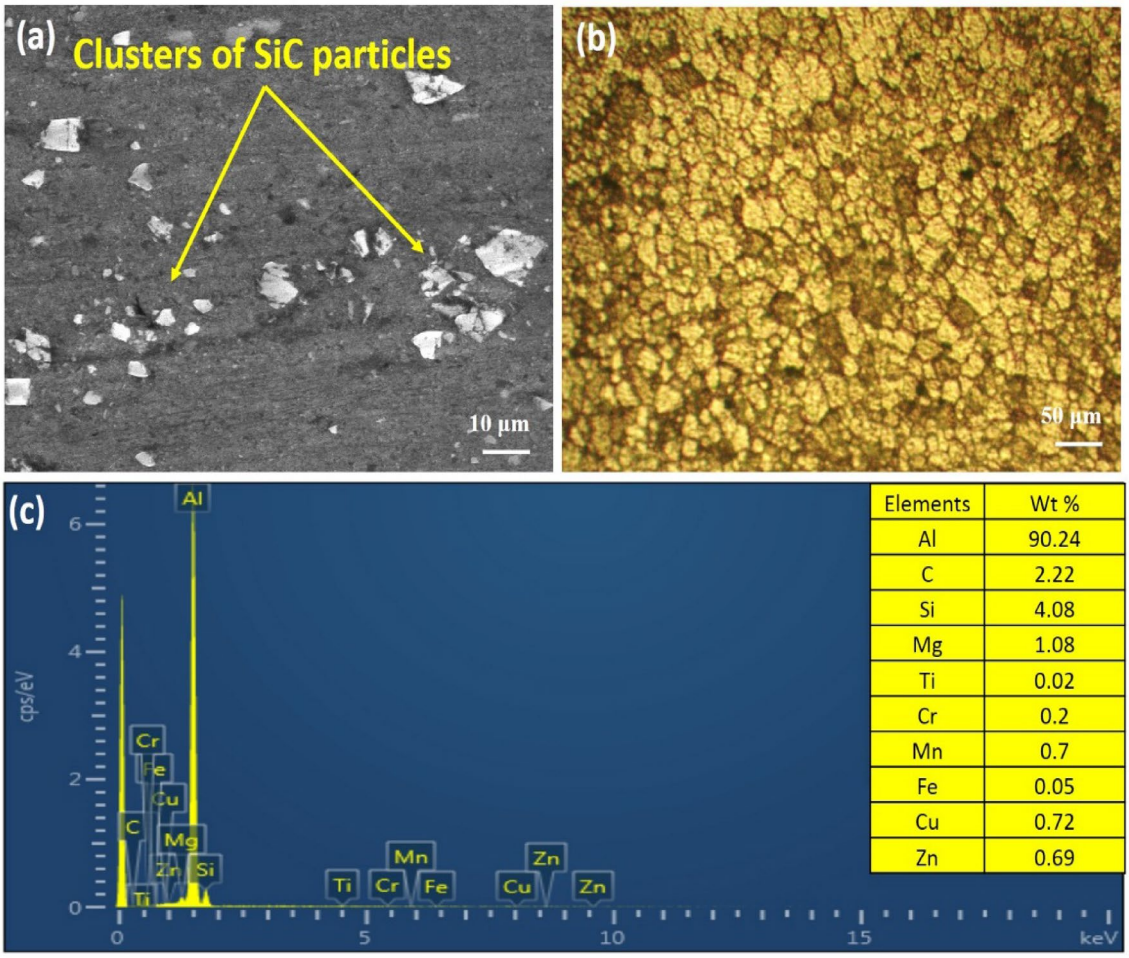


Fig. 4 **a** SEM micrograph, **b** grain structure, and **c** EDX analysis of specimen no-18

whereas specimen no-18 exhibited a minimum microhardness of 130 HV among all FSWed composite specimens. The higher hardness of specimen no-15 may be attributed to higher grain refinement via the DRX, pinning effect of uniformly dispersed hard SiCp and enhanced material mixing rule as compared to other composite joints. According to the Hall–Petch relation, smaller grains reveal higher hardness [48]. Compared to the TMAZ and HAZ, the SZ indicated the higher microhardness owing to smaller grain size caused by the pinning effect of SiCp and DRX. Lower hardness in the HAZ was observed to be caused by coarsening of strengthening precipitates, dissolution of existing clusters, and formation and coarsening of the stable phase. The processes of precipitation transformations occurring within the HAZ cannot be explicitly formulated due to susceptibility to many factors [49, 50]. In contrast, the microhardness in the HAZ was noticed less owing to coarsening of grains at high temperature and over-aging. Due to the addition of hard SiCp, the microhardness of the FSWed composite joints was also improved.

4 Development of Mathematical Model

The response function Y (UTS, %strain and microhardness) are assumed to be the function of RS (A), TS (B) and VP-SiCp (C) as represented in Eq. 1. The process parameters and the output responses were correlated by developing a second-order regression equation, and the response function Y is mathematically represented as Eq. 2.

$$Y = f(\mathrm{RS}, \mathrm{TS}, \mathrm{VP}-\mathrm{SiCp}) \tag{1}$$

The second-order polynomial equation used to represent the response surface '*Y*' is given in Eq. 2.

$$Y = b_0 + \sum b_i X_i + \sum b_{ii} X_i^2 + \sum b_{ij} X_i X_j \tag{2}$$

where *Y* is the predicted response, *bo* is the constant coefficient, *bi,* b*ii* and *bij* denote the linear coefficients, quadratic coefficients and interaction coefficients, respectively, whereas *Xi*, *Xj* are the independent variables [51, 52]. The

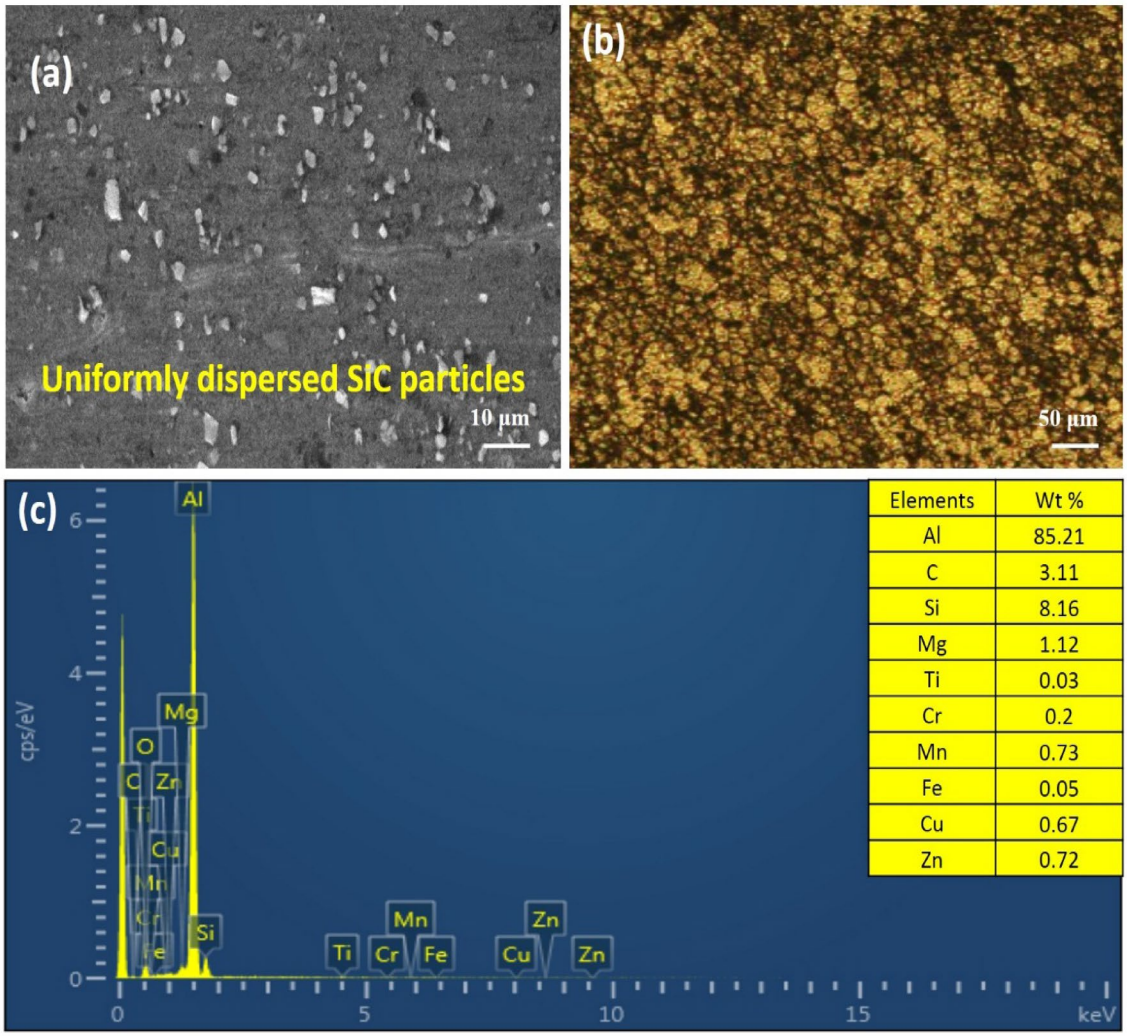


Fig. 5 **a** SEM micrograph, **b** grain structure, and **c** EDX analysis of specimen no-15

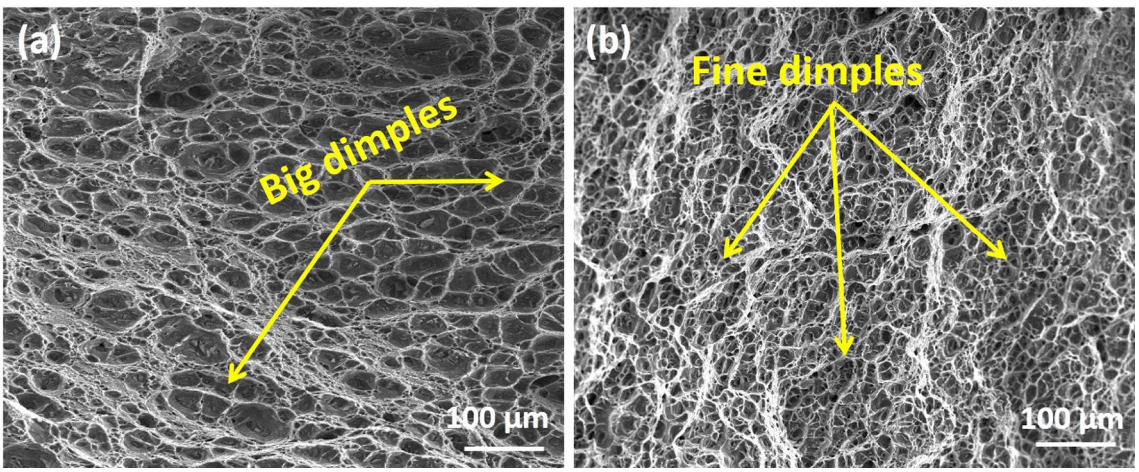


Fig. 6 Fractography SEM images; **a** specimen no-18, **b** specimen no-15

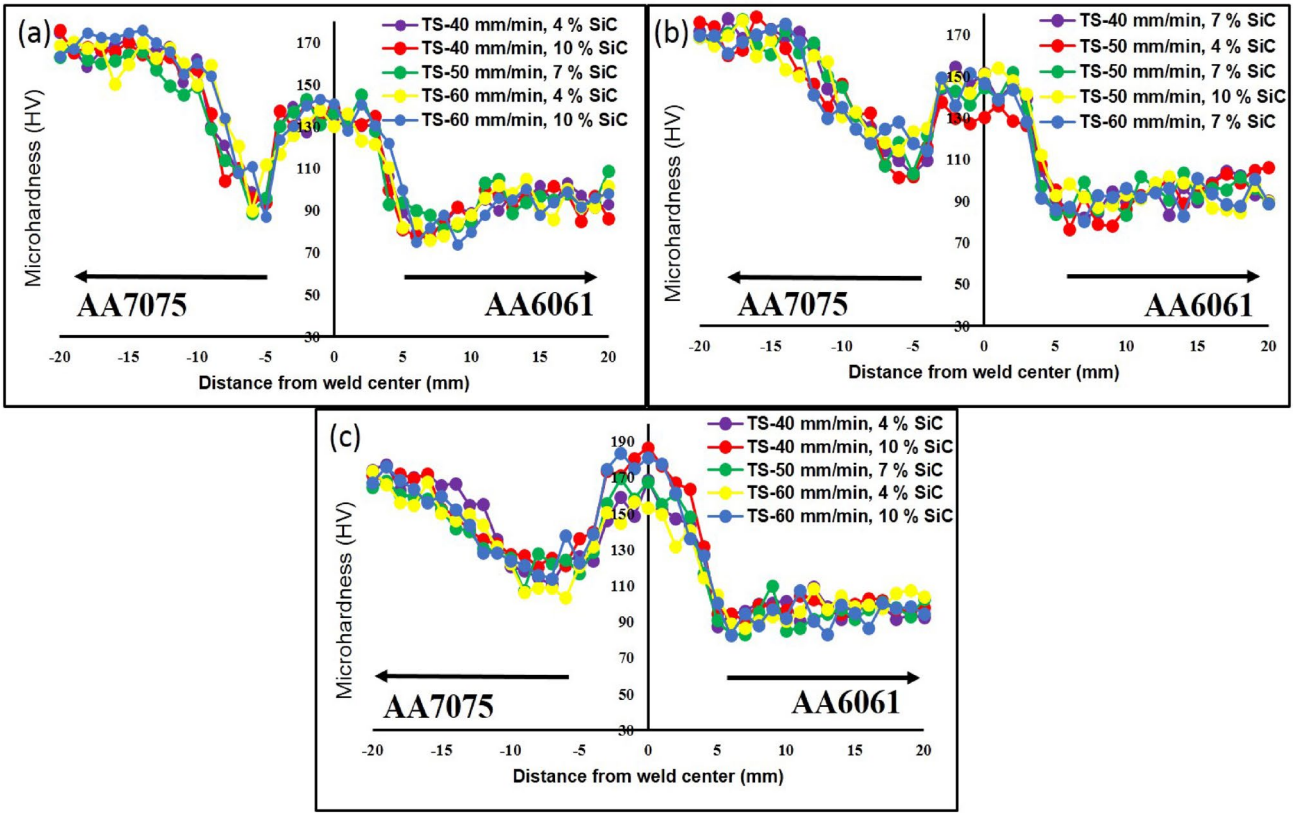


Fig. 7 Microhardness variation of the FSWed composite joints

regression model equations were developed for the response variables and written in Eqs. 3–5.

The final regression model equations developed are listed below.

$$\begin{aligned}\text{UTS} = {} & +208.03 + 25.40 * \text{RS} - 2.40 * \text{TS} + 13.00 * \text{VP - SiCp} - 3.37 * \text{RS} * \text{TS} + 3.63 * \text{RS} * \text{VP - SiCp} \\ & + 3.13 * \text{TS} * \text{VP - SiCp} + 9.68 * \text{RS}^2 + 4.68 * \text{TS}^2 - 5.32 * \text{VP - SiCp}^2\end{aligned} \tag{3}$$

$$\begin{aligned}\%\text{strain} = {} & +24.33 + 3.36 * \text{RS} + 0.28 * \text{TS} + 1.98 \\ & * \text{VP - SiCp} - 0.91 * \text{RS} * \text{TS} + 0.95 * \text{RS} \\ & * \text{VP - SiCp} - 0.51 * \text{TS} * \text{VP - SiCp} - 0.85 \\ & * \text{RS}^2 + 0.50 * \text{TS}^2 + 0.75 * \text{VP - SiCp}^2\end{aligned} \tag{4}$$

$$\begin{aligned}\text{Microhardness} = {} & +143.82 + 18.10 * \text{RS} - 2.60 * \text{TS} + 8.30 \\ & * \text{VP - SiCp} - 2.62 * \text{RS} * \text{TS} + 3.38 * \text{RS} \\ & * \text{VP - SiCp} + 1.63 * \text{TS} * \text{VP - SiCp} + 7.95 \\ & * \text{RS}^2 + 5.24 * \text{TS}^2 - 3.05 * \text{VP - SiCp}^2\end{aligned} \tag{5}$$

ANOVA is employed to identify the significance of processing variables for response variation. The results of ANOVA for UTS, %strain, and microhardness are illustrated in Tables 2, 3, and 4. Fisher's F test was performed to assess the developed models, and a 95% confidence level was used. P defines probability and should be less than 0.05 [53]. From Tables 2, 3, and 4, the F value for UTS, %strain and microhardness was found as 100.04, 306.90, and 133.96, respectively, and indicates that all three models are significant. The processing parameters (RS, TS, and VP-SiCp) are also significant, as their P value in all the mathematical models is less than 0.05. The lack of fit is not regarded as significant as the *P* value exceeds 0.1. Improved UTS, %strain and microhardness were observed with increasing volume percentage of SiC particles [54], whereas the mechanical characteristics of the FSWed composite joints were reduced with increasing TS. The *F* value developed models showed that rotating speed was the most influential parameter, volume percentage of SiCp was the second most influential parameter and transverse speed was the least influential parameter. The value of adjusted R^2 and predicted R^2 was very close for all the models, indicating high accuracy. The term "adequate

Table 2 Results of ANOVA for UTS

Sources	S.S	D.F	M.S	*F* values	*P* values	Remark
Model	9006.92	9	1000.77	100.04	< 0.0001	Significant
RS	6451.60	1	6451.60	644.94	< 0.0001	
TS	57.60	1	57.60	5.76	0.0373	
VP-SiCp	1690.00	1	1690.00	168.94	< 0.0001	
RS*TS	91.12	1	91.12	9.11	0.0129	
RS*VP-SiCp	105.13	1	105.13	10.51	0.0088	
TS*VP-SiCp	78.13	1	78.13	7.81	0.0190	
RS^2	257.78	1	257.78	25.77	0.0005	
TS^2	60.28	1	60.28	6.03	0.0340	
VP-$SiCp^2$	77.78	1	77.78	7.78	0.0192	
Residual	100.03	10	10.00			
Lack of Fit	79.20	5	15.84	3.80	0.0845	Not significant
Pure-Error	20.83	5	4.17			
Cor-Total	9106.95	19				
Std.-Dev	3.16	R^2		0.9890		
C.V. %	1.49	Pred R^2		0.9306		
Mean	212.55	Adj R^2		0.9791		
		Adeq Precision		36.48		

S.S. sum of square, *D.F.* degree of freedom, *M.S.* mean square, *F* factor F, *P* probability

Table 3 Results of ANOVA for %strain

Sources	S.S	D.F	M.S	*F* values	*P* values	Remark
Model	172.08	9	19.12	306.90	< 0.0001	Significant
RS	112.90	1	112.90	1812.13	< 0.0001	
TS	0.73	1	0.73	11.70	0.0065	
VP-SiCp	39.20	1	39.20	629.28	< 0.0001	
RS*TS	6.48	1	6.48	104.01	< 0.0001	
RS*VP-SiCp	7.22	1	7.22	115.89	< 0.0001	
TS*VP-SiCp	2.00	1	2.00	32.10	0.0002	
RS^2	1.99	1	1.99	31.89	0.0002	
TS^2	0.69	1	0.69	11.04	0.0077	
VP-$SiCp^2$	1.55	1	1.55	24.83	0.0006	
Residual	0.62	10	0.062			
Lack of Fit	0.35	5	0.071	1.32	0.3835	Not significant
Pure-Error	0.27	5	0.054			
Cor-Total	172.70	19				
Std.-Dev	0.25	R^2		0.9964		
C. V. %	1.02	Pred R^2		0.9720		
Mean	24.53	Adj R^2		0.9931		
		Adeq Precision		76.38		

S.S. sum of square, *D.F.* degree of freedom, *M.S.* mean square, *F* factor F, *P* probability

precision" is used to describe the signal-to-noise ratio. If its value exceeds 4, the model will be adequate to proceed. All three established models were given adequate precision by 36.48, 76.38, and 41.43, respectively. The value of R^2 reflects the overall variability of the models. The value of R^2 and adjusted R^2 for UTS, %strain and microhardness were found to be 0.9890 and 0.9791, 0.9964 and 0.9931, and 0.9918 and 0.9844, respectively, and their difference revealed that the models did not describe 0.99, 0.33 and 0.74% of the total variation.

Table 4 Results of ANOVA for microhardness

Sources	S.S	D.F	M.S	*F* values	*P* values	Remark
Model	4726.79	9	525.20	133.96	< 0.0001	Significant
RS	3276.10	1	3276.10	835.59	< 0.0001	
TS	67.60	1	67.60	17.24	0.0020	
VP-SiCp	688.90	1	688.90	175.71	< 0.0001	
RS*TS	55.12	1	55.12	14.06	0.0038	
RS*VP-SiCp	91.13	1	91.13	23.24	0.0007	
TS*VP-SiCp	21.13	1	21.13	5.39	0.0427	
RS^2	174.01	1	174.01	44.38	< 0.0001	
TS^2	81.82	1	81.82	20.87	0.0010	
VP-$SiCp^2$	25.51	1	25.51	6.51	0.0288	
Residual	39.21	10	3.92			
Lack of Fit	21.87	5	4.37	1.26	0.4024	Not significant
Pure-Error	17.33	5	3.47			
Cor-Total	4766.00	19				
Std.-Dev	1.98	R^2		0.9918		
C. V. %	1.33	Pred R^2		0.9498		
Mean	149.00	Adj R^2		0.9844		
		Adeq Precision		41.43		

S.S. sum of square, *D.F.* degree of freedom, *M.S.* mean square, *F* factor F, *P* probability

4.1 Influences of Processing Parameters on the Output Responses

Figure 8 shows the values of predicted vs actual output responses of various FSWed composite joints. The response variables were identified in the scattered plots to be very close to the straight lines, indicating that the model's error was distributed uniformly. There is a strong correlation between the output responses' predicted and actual values, indicating that all the regression models are adequate.

Figure 9a–f depicts the interaction influences of various processing parameters using 3D-surface plots. Examining the 3D-surface plots makes it easy to determine how different process factors influence the output responses. The key determinants in grain refining were DRX and the pinning effect of SiCp. Therefore, the output responses, including UTS, %strain and microhardness, all improved with higher RS owing to more homogeneously dispersed SiCp resulted in reduced grain size via the pinning effect.

4.2 Desirability for Multi-Optimization

RSM was employed to perform multi-response optimization in order to solve the issues of contradicting responses with single-response optimization. [55]. All output responses were given an importance of 3, and the lower and upper weights were assigned a value of 1. A set of optimal parameters with higher desirability is required in order to achieve the desired response parameters within the constraints [56]. The ramp graph's dot (Fig. 10) denoted the optimum input parameter settings for the optimum response prediction. The optimum value of RS, TS and VP-SiCp was found as 1094 rpm, 40 mm/min and 6%, respectively, whereas the value of multi-optimized UTS, %strain and microhardness was found as 247 MPa, 26.9% and 176 HV, respectively.

4.3 *Confirmatory tests*

The aim of performing confirmatory tests was to ensure that each output response could be adequately described by the developed mathematical models when the processing parameters were set to their optimum levels. The average of the results was calculated and compared to the predicted responses. Table 5 cites the confirmatory tests' results and confirms the developed models' adequacy for predicting responses.

5 Conclusions

In the current work, dissimilar composite joints of AA6061 and AA7075 were developed using SiC microparticles. The consequences of varying volume percentages of SiC particles along with RS and TS on microstructural characterization and mechanical behavior of FSWed composite joints were evaluated. The following are some of the major conclusions obtained from the findings:

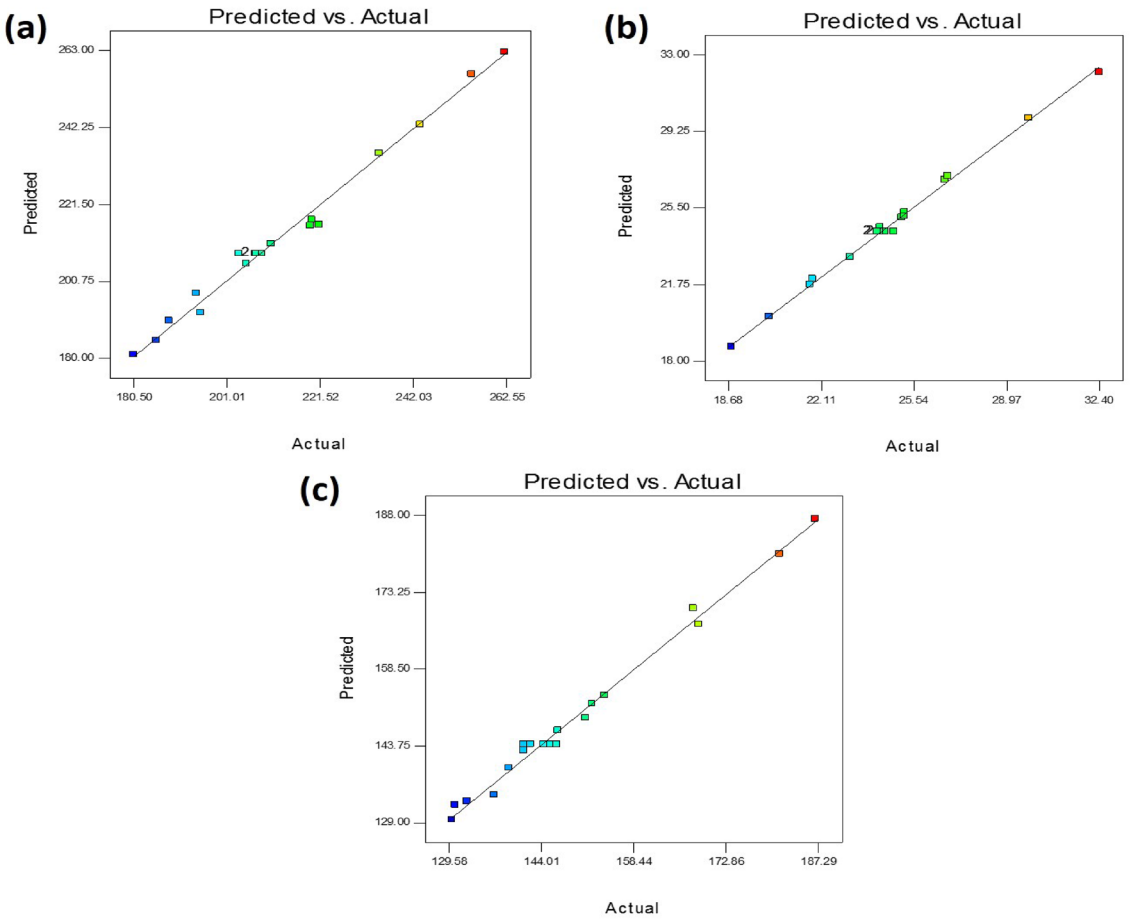


Fig. 8 **a–c** Actual vs. Predicted values plots; (**a**) UTS, (**b**) %strain, (**c**) Microhardness

(1) The significance of all three parameters in FSW lies in their ability to influence the properties of the welded joint. The developed models showed that rotating speed is the most influential parameter, volume percentage of SiCp is the second most influential parameter, and transverse speed is the least influential parameter.
(2) The higher value of rotating speed and the lower value of transverse speed led to the homogenous distribution of SiCp in the SZ, resulting in reduced grain size and enhanced mechanical properties of the FSWed composite joints.
(3) The higher UTS, %strain and microhardness were evident in the FSWed composite joints that were embedded with a higher volume percentage (10%) of SiCp.
(4) The composite joints with the highest UTS of 261 MPa, strain of 32.4%, and microhardness of 187 HV were processed at 1100 rpm, 40 mm/min, and 10% VP-SiCp.
(5) The optimum value of rotating speed, transverse speed and volume percentage of SiCp was found to be 1094 rpm, 40 mm/min and 6%, respectively. In contrast, the value of multi-optimized UTS, %strain and

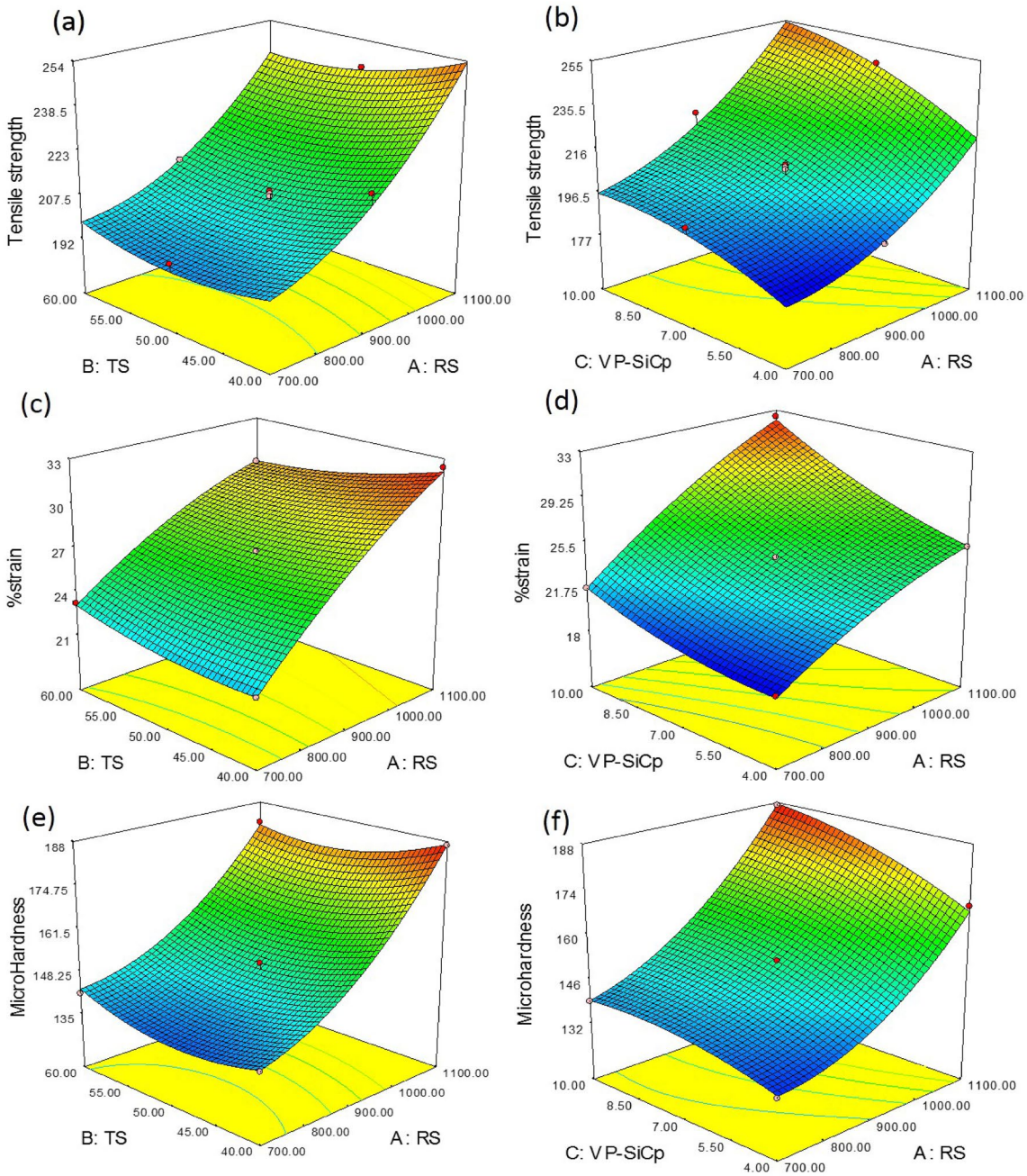


Fig. 9 **a**–**f** Surface plots; (**a**, **b**) UTS, (**c**, **d**) %strain; (**e**, **f**) Microhardness

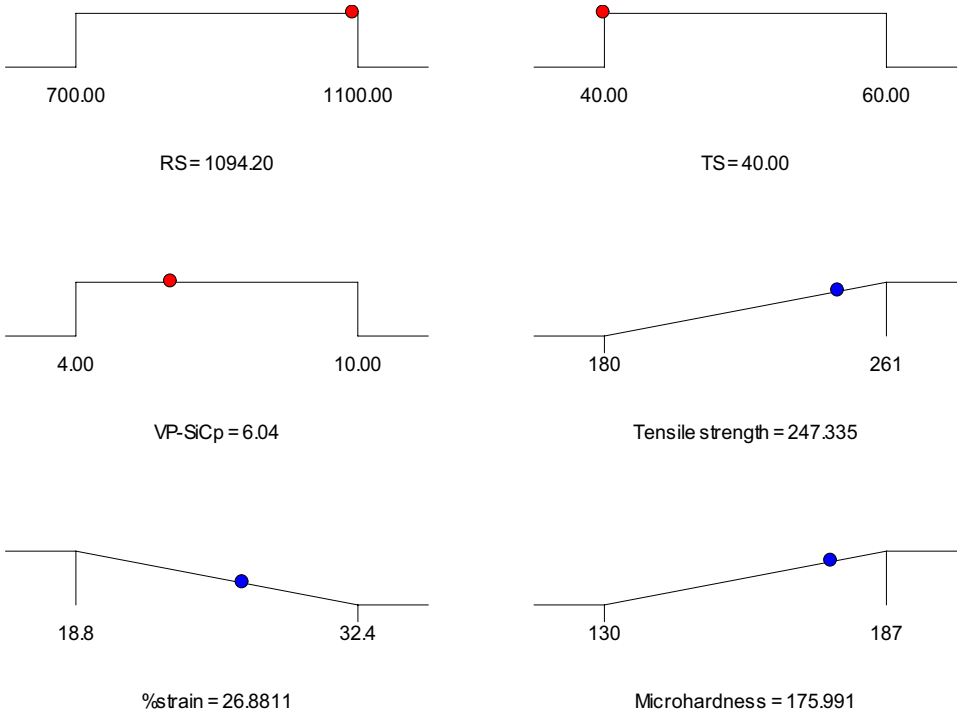


Fig. 10 Multi-optimized value of processing and response variables

Table 5 Confirmatory test

Response parameters	Optimal parameters setting	Predicted optimum results	Experimental results	% Error
UTS (MPa)	RS = 1094 rpm; TS = 40 mm/min; VP-SiCp = 6%	247	256	3.6
%strain (%)		26.9	25.6	4.5
Microhardness (HV)		176	181	2.8

microhardness was found to be 247 MPa, 26.9% and 176 HV, respectively.

Funding None.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ahmad Fauzi M N, Uday M B, Zuhailawati H, and Ismail A B, *Mater Des* **31** (2010) 670.
2. Kuryntsev S, *Materials* **15** (2022) 122. https://doi.org/10.3390/ma15010122
3. Bohm M, and Kowalski M, *Marine Struct* **71** (2020) 102739. https://doi.org/10.1016/j.marstruc.2020.102739
4. Hull D, and Clyne T, *An Introduction to Composite Materials*, Cambridge University Press, Great Britain (1996).
5. Hartley P and Hartley DE, *FSW Implementation on the Space Shuttle's External Tank, Friction Stir Welding Technology for Defense Applications*, New Orleans, LA, June 14–15, 2002.
6. Chawla K K, *Composite Materials: Science and Engineering*, Springer Science & Business Media, Heidelberg, NY (2012).
7. Kaczmar J, Pietrzak K, and Wlosinski W, *J Mater Process Technol* **106** (2000) 58.
8. Alidokht S A, Abdollah-Zadeh A, Soleymani S, and Assadi H, *Mater Des* **32** (2011) 2727.
9. Jain S, and Mishra R S, *Silicon* **14** (2022) 10741. https://doi.org/10.1007/s12633-022-01716-5
10. Barmouz M, Asadi P, Givi M B, and Taherishargh M, *Mater Sci Eng A* **528** (2011) 1740.
11. Tonelli L, Morri A, Toschi S, Shaaban M, Ammar H R, Ahmed M M Z, Ramadan R M, El-Mahallawi I, and Ceschini L, *Int J Adv Manuf Technol* **102** (2019) 3945.
12. Deore H A, Mishra J, Rao A G, Mehtani H, and Hiwarkar V D, *Surf Coat Technol* **374** (2019) 52.

13. Khodabakhshi F, Nosko M, and Gerlich A P, *Surf Coat Technol* **335** (2018) 288.
14. Uzun H, *Mater Des* **28** (2007) 1440.
15. Ceschini L, Boromei I, Minak G, Morri A, and Tarterini F, *Compos Sci Technol* **67** (2007) 605.
16. Feng A H, Xiao B L, and Ma Z Y, *Compos Sci Technol* **68** (2008) 2141.
17. Kumar K S, Murigendrappa S M, kumar H, *J Mater Res* (2019) 455.
18. Raghav G R, Nagarajan K J, Rengarajan S, Suganthi P, and Vignesh V, *Metall Res Technol* **116** (2019) 631.
19. Verma S, and Kumar V, *Proc Inst Mech Eng Part C J Mech Eng Sci* **235** (2021) 7009. https://doi.org/10.1177/09544062211005804
20. Abd Elnabi M M, El Mokadem A, and Osman T, *Int J Adv Manuf Technol* **121** (2022) 3935. https://doi.org/10.1007/s00170-022-09531-3
21. Sefene E M, and Tsegaw A A, *Int J Adv Manuf Technol* **119** (2022) 1479. https://doi.org/10.1007/s00170-021-08260-3
22. Bahemmat P, Haghpanahi M, Besharati M K, Ahsanizadeh S, and Rezaei H, *Proc Inst Mech Eng Part B J Eng Manuf* **224** (2010) 1854.
23. Çömez N, and Durmuş H, *ASME J Eng Mater Technol* **141** (2019) 031002.
24. Aherwar A, Patnaik A, and Pruncu C I, *J Mater Res Technol* **9** (2020) 9882. https://doi.org/10.1016/j.jmrt.2020.07.003
25. Rathee S, Maheshwari S, Siddiquee A N, and Srivastava M, *Def Technol* **13** (2017) 86.
26. Rathee S, Maheshwari S, Siddiquee A N, and Srivastava M, *Mater Today Proc* **4** (2017) 452.
27. Godhani P S, Patel V V, Vora J J, Chaudhary N D, and Banka R, *Effect Friction Stir Welding of Aluminum Alloys AA6061/AA7075: Temperature Measurement, Microstructure, and Mechanical Properties*. Springer Singapore.
28. Cole E G, Fehrenbacher A, Duffie N A, and Zinn M R, *Weld temperature effects during friction stir welding of dissimilar aluminum alloys 6061-t6 and 7075-t6*, (2014), pp 643–652.
29. Mehdi H, and Mishra R S, *Adv Mater Process Technol* (2020). https://doi.org/10.1080/2374068X.2020.1829952
30. Mehdi H, and Mishra R S, *J Mater Eng Perform* **30** (2021) 1926.
31. Srinivasu R, Rao A S, Reddy G M, and Rao K S, *Def Technol* **11** (2015) 140.
32. Mehdi H, and Mishra R, *Proc Inst Mech Eng Part L J Mater Design Appl* **235** (2021) 2531. https://doi.org/10.1177/14644207211007882
33. Mehdi H, and Mishra R S, *J Adhes Sci Technol* (2021). https://doi.org/10.1080/01694243.2021.1964846
34. Yadav D, and Bauri R, *Mater Sci Eng A* **528** (2011) 1326.
35. Salah A N, Mabuwa S, Mehdi H, Msomi V, Kaddami M, and Mohapatra P, *Silicon* (2022). https://doi.org/10.1007/s12633-022-01717-4
36. Chen X G, Silva M D, Gougeon P, and Georges L S, *Mater Sci Eng A* **518** (2009) 174.
37. Palanivel R, Dinaharan I, Laubscher R F, and Paulo D J, *Mater Des* **106** (2016) 195.
38. Srivastava M, Rathee S, Siddiquee A N, and Maheshwari S, *Silicon* **11** (2019) 2149.
39. Azizieh M, Kokabi A H, and Abachi P, *Mater Design* **32** (2011) 2034.
40. Chang C I, Lee C J, and Huang J C, *Scr Mater* **51** (2004) 509.
41. Mehdi H, and Mishra R S, *Trans Indian Inst Met* **73** (2020) 1773.
42. Lloyd D J, *Int Mater Rev* (1994) 39.
43. Salehi M, Saadatmand M, and Aghazadeh Mohandesi J, *Trans Nonferrous Metals Soci China* **22** (2012) 1055.
44. Nait Salah A, Mehdi H, Mehmood A, Hashmi A W, Malla C, and Kumar Ravi, *Mater Today Proc* (2021). https://doi.org/10.1016/j.matpr.2021.10.288
45. Pandey C, Saini N, Mahapatra M M, and Kumar P, *Eng Fail Analy* **71** (2017) 131.
46. Mehdi H, and Mishra R S, *SN Appl Sci* **1** (2019) 712.
47. Xu W, Liu J, Luan G, and Dong C, *Mater Design* **30** (2009) 1886.
48. Ebrahimzadeh V, Paidar M, Safarkhanian M A, and Ojo O O, *Int J Adv Manuf Technol* **96** (2018) 1237.
49. Jain S, Mishra R S, Effect of Al2O3 nanoparticles on microstructure and mechanical properties of friction stir-welded dissimilar aluminum alloys AA7075-T6 and AA6061-T6, 2021, https://doi.org/10.1177/09544089211065534.
50. Rejil C M, Dinaharan I, Vijay S J, and Murugan N, *Mater Sci Eng A* **552** (2012) 336.
51. Yiew T, Ling E, Fadhil M, Rezania S, Aminudin E, Chelliapan S, Abdul A, and Park J, *J Environ Manag* **228** (2018) 13. https://doi.org/10.1016/j.jenvman.2018.09.008
52. Sadeghian M, Sadeghi M, Hesampour M, and Moheb A, *Desalin Water Treat* (2014). https://doi.org/10.1080/19443994.2014.919607
53. Puri A B, and Bhattacharyya B, *Int J Adv Manuf Technol* **25** (2005) 301.
54. Patel V V, Badheka V, and Kumar A, *Mater Manuf Proce* **31** (2016) 1573.
55. Ghangas G, and Singhal S, *Mater Res Express* **6** (2018) 026553.
56. Shanavas S, and Dhas J E R, *Trans Nonferrous Metals Soc China* **27** (2017) 2334.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

This article has been accepted for publication in IEEE Transactions on Components, Packaging and Manufacturing Technology. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/TCPMT.2023.3304490

MANUSCRIPT ID TCPMT-2022-342

Partially Omnidirectional and Circularly Polarized MIMO Antenna covering sub-6 GHz Band for 5G Fast Plan

Monika Kaushik, Jasdeep K. Dhanoa, *Senior Member, IEEE,* and Mukesh K. Khandelwal, *Senior Member, IEEE*

***Abstract*— A two-port MIMO antenna is designed, manufactured, and analyzed, which covers the whole sub-6 GHz band ranging from 0.97 - 6.19 GHz with an integrated circularly polarized wideband centred at 3.5 GHz. The proposed work overcomes the challenges of (i) large size of antennas at *L*-band, (ii) circular polarization for wideband, (iii) channel capacity for 5G, and (iv) diversity in MIMO, and (v) minimum mutual coupling with inbuilt isolator. A very small antenna size of $0.113\lambda_0 \times 0.129\lambda_0$ is obtained after incorporating several techniques, including slots, stubs, defected ground, and parasitic elements. The structure is simple and contains an in-built isolator that provides more than 23 dB isolation between both ports. Partial ground with additional stubs helps to maintain high isolation between both ports within the whole operating band. The proposed antenna also improves the channel capacity and provides the minimum channel capacity loss (CCL) of about 0.12 bits/s/Hz. Various diversity characteristics (e.g. ECC, TARC, DG) are analyzed and found suitable within the required range. The proposed antenna radiates the omnidirectional pattern at the *L1* (1.57 GHz) and *L2* (1.22 GHz) bands of the Global Positioning System (GPS). Other than *L1* and *L2* bands, the proposed antenna also offers wide beamwidth for the remaining bands. Finally, an antenna model is manufactured to validate the findings through experimental measurements.**

***Index Terms*— Sub-6 GHz, 5G Fast Plan, MIMO Antenna, circularly polarized wideband, antenna packaging,**

I. INTRODUCTION

THE Sub-6 GHz band has been proposed for both types of licensed and unlicensed applications of fifth-generation (5G) networks by the Federal Communications Commission (FCC), Global System for Mobile Communications Association (GSMA), and 3rd Generation Partnership Project (3GPP) [1-6]. Moreover, the 3.5 GHz band within the sub-6 GHz range has been suggested for fast 5G networks by FCC [3-4], GSMA [5], and 3GPP [6]. Recently in April 2021, GSMA proposed a 3.5 GHz band (3.3 GHz to 4.2 GHz) for the implementation of 5G services globally [5]. Secondly, capacity and coverage of the band are two main factors for wireless applications. The band's capacity can be enhanced by using the multiple-in multiple-out (MIMO) technique [6]. Further, coverage can be improved by controlling the beamwidth and polarization of the antenna. Circularly polarized waves have minimum propagation losses and thus cover a maximum range [7]. However, achieving circular polarization (CP) within a wide range of frequency bands is challenging. Therefore, to address all the issues mentioned above of 5G, the proposed work is presented here, which covers the whole sub-6 GHz band ranging from 0.97 – 6.19 GHz, including a wide circularly polarized band of 1.43 GHz centred at 3.5 GHz, ranging from 2.7 GHz to 4.13 GHz. Here, a two-port MIMO is proposed, enhancing the channel capacity throughout the operating band. Therefore, one antenna can be used for multiple Sub-6 GHz wireless applications simultaneously.

Several MIMO microstrip antennas have been developed for wideband MIMO applications [8-12]. However, most of them are reported for UWB (ultrawideband) range (3.1-10.6 GHz). Because the 5G communication applications of lower frequencies are to be implemented within the sub-6 GHz band only, the upper band (from 6GHz to 10.6GHz) of the UWB antenna creates interference if it is used for such 5G applications. These unwanted interfering bands limit the performance of any wideband MIMO sub-6 GHz system. Although, few MIMO antennas have been reported for 5G applications under sub-6 GHz bands [13-15]. Moreover, few wideband antennas are also noted for lower 5G band applications [16-17]. Metamaterials and Metasurfaces have been used in MIMO antennas to improve the mutual coupling and other parameters [18-27]. However, there is still scope for improvement and advancement in circular polarization, mutual coupling, CCL, etc. Additionally, few antennas are reported that partially serve the sub-6 GHz band [28-30]. However, it is tough to find such a wideband MIMO antenna in the literature that covers the sub-6 GHz band, including L-band. The antenna proposed in this study not only bridges this identified gap of literature but also is integrated with a wide circularly polarized band centered at the most significant frequency for 5G applications.

The presented work proposes a wideband two-port MIMO antenna that covers the whole sub-6 GHz band starting from

Manuscript received on Aug 21, 2022, and revised on Nov 20, 2022, and Aug 05, 2023..

M. Kaushik is with Electronics & Communication Department, Indira Gandhi Delhi Technical University for Women, New Delhi, India-110006 and Bhagwan Parshuram Institute of Technology, New Delhi, India-110089. (e-mail: kaushikmonika52@gmail.com).

J. K. Dhanoa is with Electronics & Communication Department, Indira Gandhi Delhi Technical University for Women, New Delhi, India-110006 (e-mail: jasdeepkaur@yahoo.com).

M. K. Khandelwal is with School of Engineering, Jawaharlal Nehru University, New Delhi, India-110067. (corresponding author; phone: +91-9015580922; e-mail: mukesh.khandelwal89@gmail.com).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:41:21 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Components, Packaging and Manufacturing Technology. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/TCPMT.2023.3304490

0.97 GHz. The operating band is integrated with a circularly polarized band of 1.43 GHz wide bandwidth and centred at the most significant frequency for 5G applications, i.e., 3.5 GHz. Additionally, the proposed antenna radiates the omnidirectional pattern for the *L1* (1.57 GHz) and *L2* (1.22 GHz) bands of the Global Positioning System (GPS). Furthermore, the proposed antenna also consists of an in-built isolator to provide the minimum mutual coupling. Moreover, the proposed design is simple and easy to construct with minimum fabrication losses. A detailed parametric study of the proposed structure is presented, and various diversity parameters are analyzed in this work.

II. Antenna Design, Mechanism, and Analysis

This section describes the proposed antenna's design concept, evolution, and mechanism. The construction of the proposed antenna (*Ant-6*) is achieved through five intermediate antenna models named *Ant-1* to *Ant-5*, and the final structure is represented as *Ant-6* as depicted in Fig 1 (a). It is basically a two-port microstrip MIMO antenna constructed on a common ground plane. The presented antenna is designed and fabricated on a Fr-4 Epoxy substrate. The relative permittivity and loss tangent of the used substrate is 4.4 and 0.02, respectively.

The evolution steps are shown in Fig 1 (b). *Ant-1* is also achieved through three stages named I to III in the left corner inset of Fig 1 (b).The values of all the dimensions are (all in mm): L=35, W=40, h=1.57, L_1=10, L_2=4, L_3=1, L_4=1, L_5=6, L_6=2, L_f=10, W_1=10, W_2=6, W_3=2, W_4=2, W_5=8, W_6=4, W_f=3.01, W_g=6, W_d=8, G_d=4, D_1=1, D_2=1.1, H_1=15, H_2=15.9, G_1=14.5, G_2=2, St_x=3, St_y=24, $S1_x$=0.1, $S1_y$=4.4.

The basic idea of design and step-by-step construction of the final antenna design through five intermediate structures are described in the following subsections:

A. Design of basic antenna structure (Ant-1)

This study starts with a basic structure consisting of two identical microstrip line-fed patches supported by a common partial ground plane and as shown in the *I*-inset of the left corner of Fig 1 (b). The antenna structure is fed with a microstrip line of 50Ω characteristic impedance. Basically, it is a microstrip line fed monopole antenna supported by a semiground plane. The width W_f and length L_f of microstrip line are calculated by well-known mathematical equations (1) to (3) [31].

$$W_f = \frac{2h}{\pi}\left\{[B-1-ln(2B-1)] + \frac{\varepsilon_r-1}{2\varepsilon_r}\left[ln(B-1)+0.39-\frac{0.61}{\varepsilon_r}\right]\right\} \quad (1)$$

where, B is a constant having value equivalent to $377\pi/100\sqrt{\varepsilon_r}$

$$\varepsilon_{eff} = \frac{\varepsilon_r+1}{2} + \frac{\varepsilon_r-1}{2}\left(\sqrt{1+\frac{12h}{W_f}}\right)^{-1} \quad (2)$$

$$L_f = c \times \left(4f_d\sqrt{\varepsilon_{eff}}\right)^{-1} \quad (3)$$

The length L_1 and width W_1 of the patch is same and taken as

$$L_1 = W_1 = \frac{\lambda_{eff}}{4} = \frac{c}{4 \times f_d\sqrt{\varepsilon'_{eff}}} \approx 10\ mm \quad (4)$$

where c and f_d are the speed of light in free space and design frequency, respectively. ε_r is the relative permittivity of the substrate. Design frequency f_d is taken as 3.8 GHz. Effective permittivity ε'_{eff} is calculated as per eq (2).

After stage-I, a rectangular slot of length L_5 and width W_5 is embedded at the centre of the square patch for stage-II of Fig 1 (b). For further stage-III, a rectangular copper ring of outer dimensions $L_2 \times W_2$ is inserted within the recently constructed slot. Finally, one more rectangular copper patch of dimension $L_3 \times W_3$ is inserted at the common centre of this patch and newly made structure is referred as *Ant-1*. The *Ant-1* is called here as a reference antenna in this study which is established through four significant steps (stage-I to stage-IV), as shown in the left corner of Fig 1 (b). Stage-IV represents *Ant-1* itself. The S_{11} and S_{12} characteristics from stage-I to stage-IV is shown in Fig 1 (c), and Fig 1 (d), respectively.

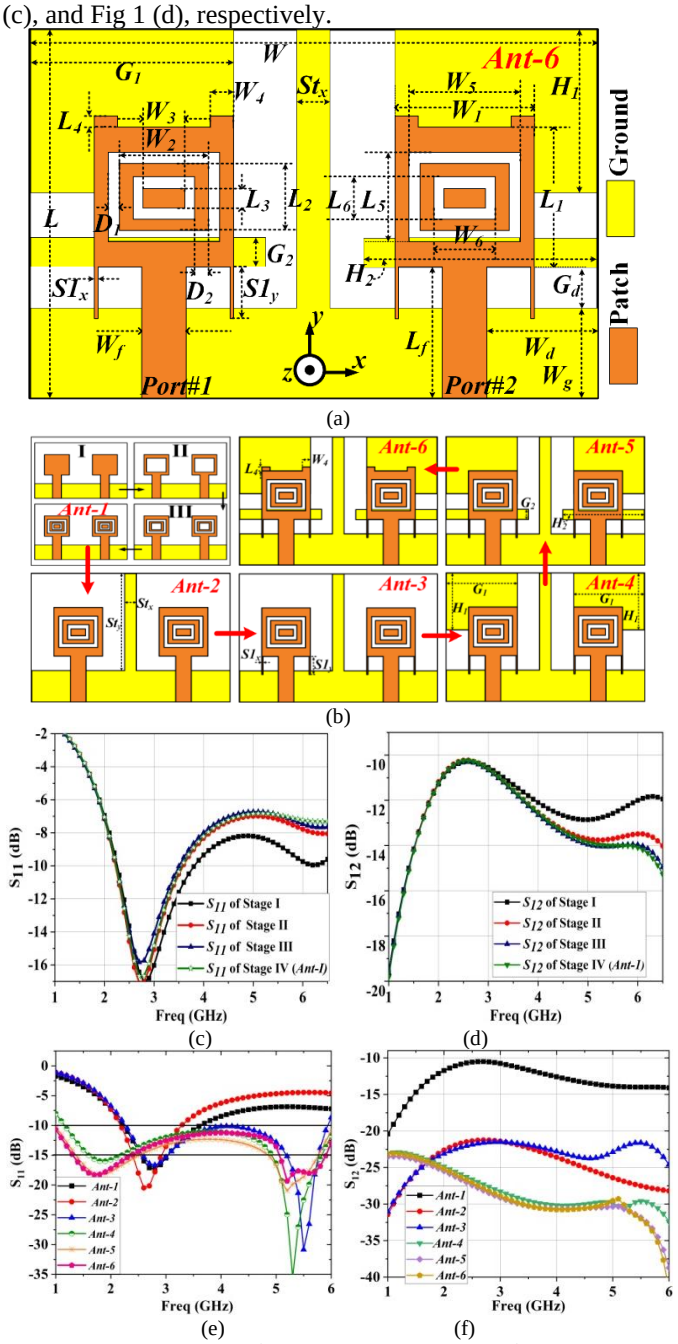


Fig. 1. Schematic designs of the proposed antenna; (a) final structure, (b) evolutionary steps, (c) S_{11} characteristics of stage-I to stage-IV, (d) S_{12} characteristics of stage-I to stage-IV, (e) S_{11} characteristics of *Ant-1* to *Ant-6*, (f) S_{12} characteristics of *Ant-1* to *Ant-6*

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:41:21 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Components, Packaging and Manufacturing Technology. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/TCPMT.2023.3304490

Basically, *Ant-1* is a microstripline fed slotted square patch monopole antenna supported by a semi-ground plane. Semiground plane is useful to achieve a wide impedance bandwidth as it sends back the reflected waves coming towards the port, thus it improves the impedance matching over a wide range of frequency. The gap G_d between the upper edge of the semiground plane and lower edge of patch, plays a vital role in controlling the return loss level within the operating band. Parametric analysis of ground's length W_g is observed and shown in Fig 2 (a) and (b). Length W_g is basically changing the gap G_d. Initially, the design frequency is considered 3.8 GHz; however, miniaturization is achieved in the later stages of this study. The transmission line's width is set to 3.01 mm corresponding to 50-Ω characteristic impedance to get minimum reflection loss.

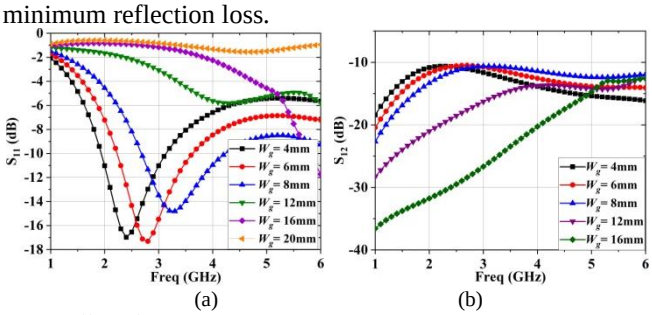


Fig. 2. Effect of length, W_g; (a) S_{11} characteristics, and (b) S_{12} characteristics

S_{11} and S_{12} characteristics of stage-*I* to *Ant-1* are shown in Fig 1 (c) and (d), respectively. *Ant-1* offers good return loss in the frequency range of 2.2 GHz to 3.6 GHz, however, both ports are closely placed, thus, poor isolation is noticed between them. Both ports are separated by a small distance to maintain compactness; however, the mutual coupling will be suppressed in further steps of this study even with this small separation distance.

B. Analysis and construction from Ant-2 to Ant-6

Fig 1 (c) and (d) are evident that *Ant-1* is not covering the whole sub-6 GHz band and also have low isolation between both ports. Therefore, the reference antenna (*Ant-1*) is modified to further stages in the form of *Ant-2* to *Ant-6*. For *Ant-2*, a vertical rectangular strip of dimensions $St_x \times St_y$ is attached to the semi-ground plane as shown in the schematic of *Ant-2* in Fig 1 (b). This vertical strip on the ground plane perturbs the coupling energy travelling between both ports, thus suppressing the mutual coupling. *Ant-2* offers a better S_{11} and S_{12} than the reference antenna (*Ant-1*); however, it does not cover the complete sub-6 GHz band. Therefore, further work is carried out to enhance the operating bandwidth with minimum mutual coupling and to achieve circular polarization.

Therefore, for further extension, two symmetrical rectangular stubs of dimension $S1_x \times S1_y$ are attached at the lower edges of both patches, as shown in the schematic of *Ant-3* in Fig 1 (b). We are calling these stubs as patch stubs here. The following three types of additional capacitance and inductance are introduced after embedding a pair of identical stubs; (i) a capacitance between the inner edges of these patch stubs, (ii) a capacitance between ground and patch stubs, and (iii) self and mutual inductance due to these patch stubs. These additional capacitances and inductance affect the resonance and improve the return loss characteristics for a wider impedance bandwidth.

The effects of length $S1_y$ and width $S1_x$ of the patch stub on S_{11} of *Ant-3* are shown in Fig 3 (a) and (b), respectively. The maximum bandwidth for *Ant-3* is obtained at $S1_y = 4.4$ mm and $S1_x = 0.1$ mm. *Ant-3* offers impedance bandwidth from 2.3 GHz to 5.98 GHz, in which mutual coupling is below -22 dB.

Ant-3 still doesn't cover the whole sub-6 GHz band; moreover, circular polarization is still to be achieved. Therefore, to improve the impedance bandwidth furthermore, two identical rectangular conducting strips of size $H_1 \times G_1$ are embedded on the reference plane to construct the *Ant-4*, as shown in the schematic of *Ant-4* of Fig 1 (b). Newly added conducting portions enhance the effective aperture of the antenna and improve the impedance matching corresponding to the L-band also. Thus, *Ant-4* provides the impedance bandwidth from 1.45 GHz to 6.2 GHz at H_1=15mm and G_1=14.5mm. The mutual coupling is below -22 dB for the complete working frequency. The effects of width G_1 and height H_1 of these conducting patches on the return loss of *Ant-4* are plotted in Fig 3 (c) and (d), whereas their effects on the axial ratio (AR) of *Ant-4* are presented in Fig 4 (c) and (d), respectively.

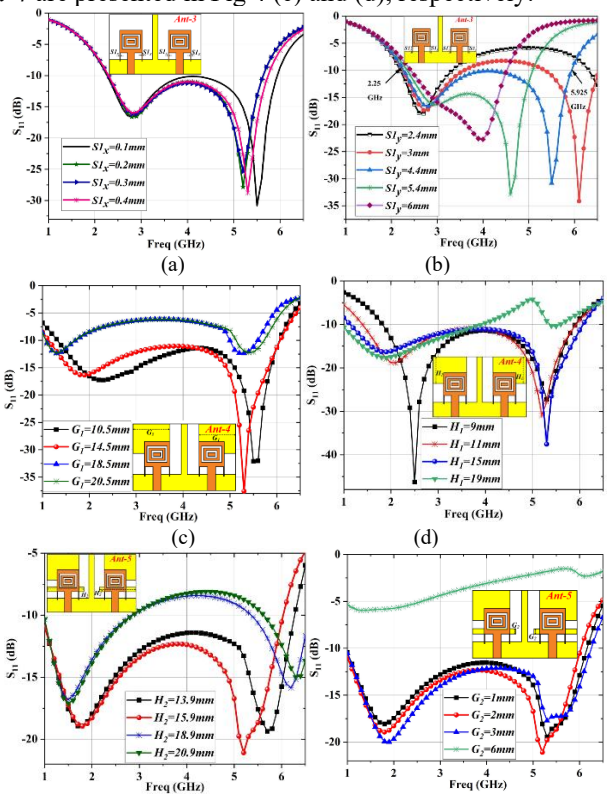


Fig. 3. Parametric S_{11} analysis, the effect of; (a) length *Sly* on *Ant-3*, (b) width *Slx* on *Ant-3*, (c) width G_1 on *Ant-4*, (d) length H_1 on *Ant-4*, (e) length H_2 on *Ant-5*, (f) width G_2 on *Ant-5*

In further analysis steps, two symmetric conducting rectangular strips of $G_2 \times H_2$ dimensions are embedded on the ground plane, which helps to enhance the range of impedance bandwidth. The new structure with integrated rectangular strips in the ground is named *Ant-5,* operating from 1 GHz to 6 GHz.

The effects of variation of length H_2 and width G_2 of these rectangular strips on the impedance bandwidth of *Ant-5* are shown in Fig 3 (e) and (f), whereas their effects on AR of *Ant-5* are presented in Fig 4 (e) and (f), respectively. The parametric

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:41:21 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Components, Packaging and Manufacturing Technology. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/TCPMT.2023.3304490

analysis confirms that maximum bandwidth is achieved at strip length H_2=15.9 mm and strip width G_2=2 mm.

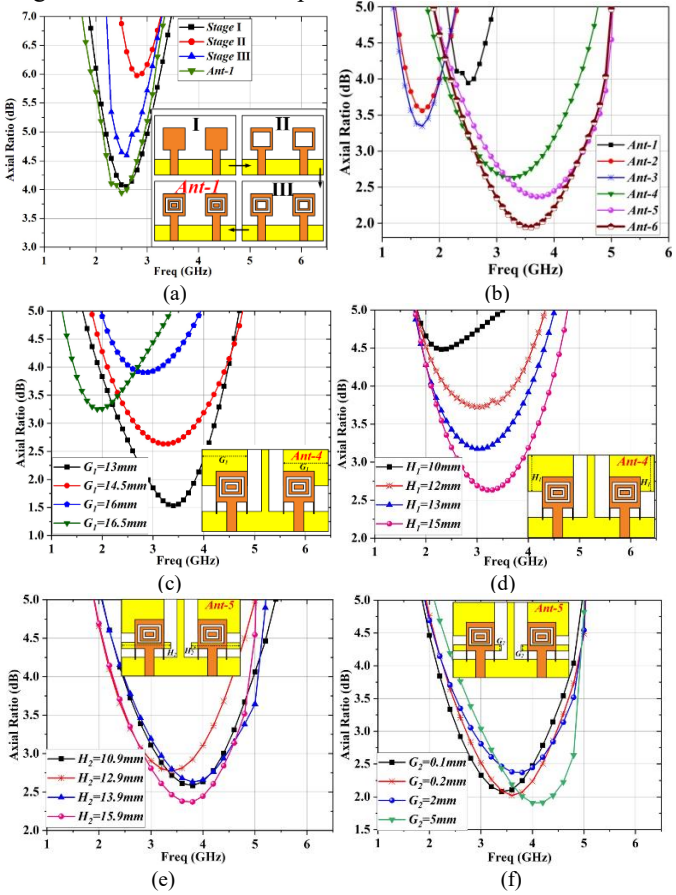


Fig. 4. Parametric axial ratio analysis of; (a) stage-I to *Ant-1*, (b) *Ant-1* to *Ant-6*, (c) effect of width G_1 on *Ant-4*, (d) effect of length H_1 on *Ant-4*, (e) effect of length H_2 on *Ant-5*, (f) effect of width G_2 on *Ant-5*

Finally, two rectangular strips of dimensions $W_4 \times L_4$ are added at the upper corners of both the radiating elements, as shown in the schematic of *Ant-6* in Fig 1. The optimized distance between these two rectangular strips is 6 mm. The introduction of strips in the patch further improves the reflection coefficient, as shown in the analysis of S-parameters given in Fig 1 (e) and (f). For this antenna, the mutual coupling is below -22 dB throughout the whole operating band from 0.97 GHz to 6.19 GHz. Therefore, the proposed antenna (*Ant-6*) covers the whole sub-6 GHz band starting from 0.97 GHz to 6.19 GHz. Moreover, *Ant-6* provides a good axial ratio at desired frequency i.e. 3.5 GHz. A comparison of S_{11}, and S_{12} of *Ant-1* to *Ant-6* is given in Fig 1 (e), and (f), respectively, whereas their AR comparison is shown in Fig 4 (b). *Ant-6* is the final product of this study. A prototype model is fabricated for this proposed antenna whose photograph is given in Fig 5 (a). Simulated and measured S-parameters results of *Ant-6* are shown in Fig 5 (b). *Ant-6* provides suppressed mutual coupling with a minimum value of -23 dB within the working frequency. About 74.47% [$\{(f_d\text{-}f_l)/f_d\}\times100$] of miniaturization is achieved corresponding to the design frequency (f_d=3.8 GHz), where f_l is the lowest working frequency. Ansys High-Frequency Structure Simulator (HFSS) is used for simulation work. Experimental results of the fabricated antenna are measured using a two-port Anritsu MS2028 VNA (Vector Network Analyzer).

The dimensions of the *Ant-6* are $L\times W\times h$, where L=0.113λ_0, W=0.129λ_0, and h=1.57mm represent the substrate's length, width, and height, respectively. λ_0 is taken corresponding to the lowest working frequency, i.e. f_l=0.97 GHz.

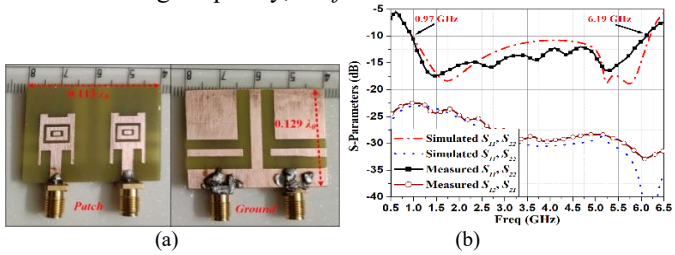
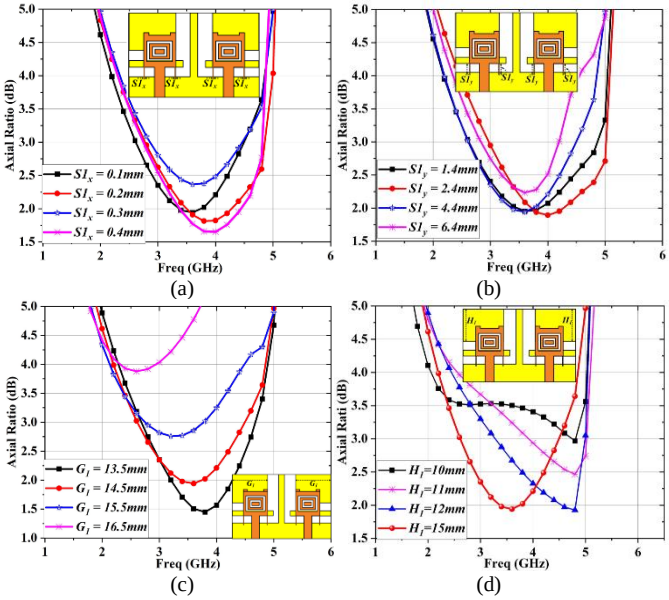


Fig. 5. (a) Fabricated prototype, (b) simulated and measured S-parameters of *Ant-6*

C. Axial Ratio (AR) Analysis

The parametric effects of different stages and dimensions on the axial ratios of intermediate structures are shown in Fig 3. The AR of stage-*I* to *Ant-1* is shown in Fig 4 (a) whereas a comparison of the AR of *Ant-1* to *Ant-6* is presented in Fig 4 (b). *Ant-1* to *Ant-3* have close results however, they are still above the satisfactory limit of 3dB. Therefore, further structures were evolved to achieve good AR at 3.5 GHz with a targeted coverage range. A parametric study of AR of *Ant-4* and *Ant-5* is given in Fig 4 (c) to (f). It is evident that defected ground structure of the proposed antenna is contributing to modeling the AR of the structure, however, it is also supported by the main radiating patch's structure. The effect of the variation of dimensions of *Ant-6* on AR is presented in Fig 6. It is observed that the circular polarization in *Ant-6* is credited due to the combined effect of multiple rectangular slots in radiating elements, rectangular strips in the ground planes, and rectangular strips at the upper corners of both the radiating elements. Fig 6 shows the measured and simulated AR of *Ant-6*, providing a wide axial ratio 3dB bandwidth of 1.68 GHz, centred at 3.55 GHz and ranging from 2.7 GHz to 4.38 GHz.



The circular polarization of the proposed antenna is also justified through the current distribution shown in Fig 8. The surface current density vectors at 3.55 GHz frequency are represented at 0º, 45º, 90º, 135º, 180º, 225º, 270º, 315º, and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:41:21 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Components, Packaging and Manufacturing Technology. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/TCPMT.2023.3304490

360° phases. The left-hand circularly polarized behaviour (LHCP) is observed from the current distribution.

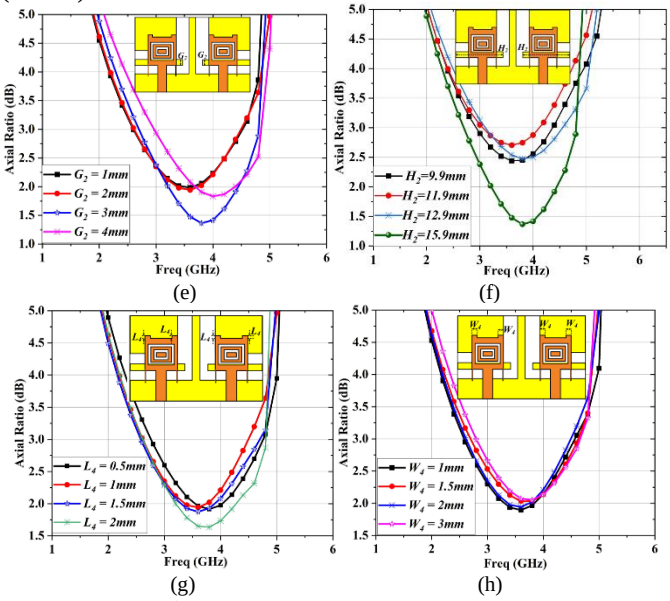


Fig. 6. Axial ratio of *Ant-6* with variation in; (a) $S1_x$, (b) $S1_y$, (c) G_1, (d) H_1, (e) G_2, (f) H_2, (g) L_4, (h) W_4

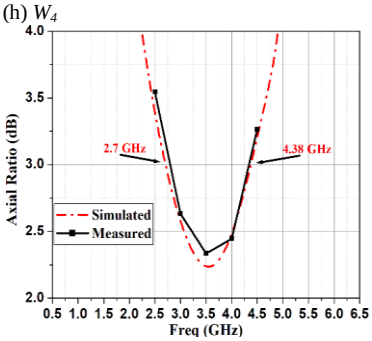


Fig. 7. Simulated and measured axial ratio of *Ant-6*

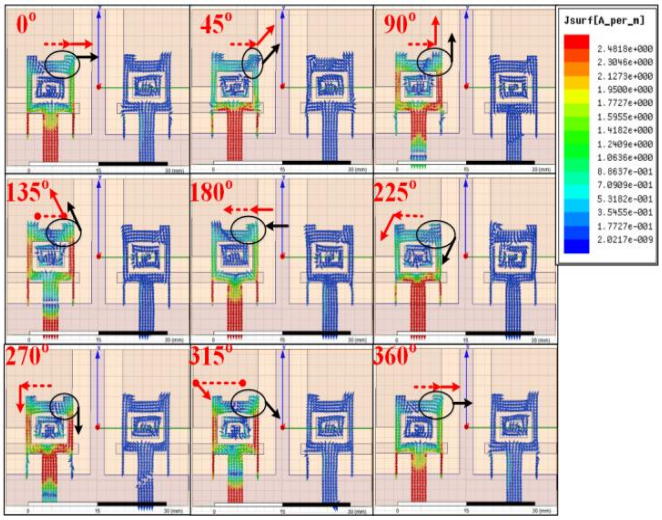


Fig. 8. Current distribution at 3.55 GHz for the different phases of the input signal (phases are mentioned in degree).

D. Radiation Pattern

The normalized radiation patterns of the *Ant-6* are shown in Fig 8. The radiation patterns are recorded in all three major planes (i.e. XY, XZ, and YZ planes) to provide better visualization and understanding of the radiation characteristics. Patterns are observed at 1.22, 1.57, 2.4, 3.55, 5.1, and 5.8 GHz frequencies, covering the almost wide band and representing the major applications for GPS, WLAN, Wi-MAX, and ISM (industrial, scientific, and medical) bands, etc. Frequencies 1.55 GHz as L1 and 1.22 GHz as L2 are used for GPS satellites. From the radiation patterns shown in the first two figures of Fig 8, it is observed that the proposed antenna is omnidirectional in YZ and XZ planes because patterns are beyond the -3dB circle in both planes at these two frequencies. Moreover, it is directional with broad HPBW (half power beam width) in the XY plane also at both frequencies. Therefore, the proposed antenna *Ant-6* is a suitable candidate for GPS applications in the most preferable L1 and L2 bands.

Further, it is noticed for 2.4 GHz frequency that *Ant-6* is completely omnidirectional in the YZ plane and almost omnidirectional in the XZ plane (except 55° to 95°). The 2.4 GHz band is the most useful band for modern wireless applications. For this band, the antenna can be used for ISM (industrial, scientific and medical), WLAN, and WiMAX. The achievement of the omnidirectional pattern at this frequency makes the proposed antenna most prominent for these applications. It is also observed that *Ant-6* provides broad HPBW throughout the operating bands in all major planes. The radiation pattern of the proposed antenna varies with frequency. There may be several reasons, and out of them, one is that the antenna is composed of several resonators, who are contributing to achieve a wide-band. Further, FR-4 epoxy is a lossy substrate whose losses are not stable with frequency.

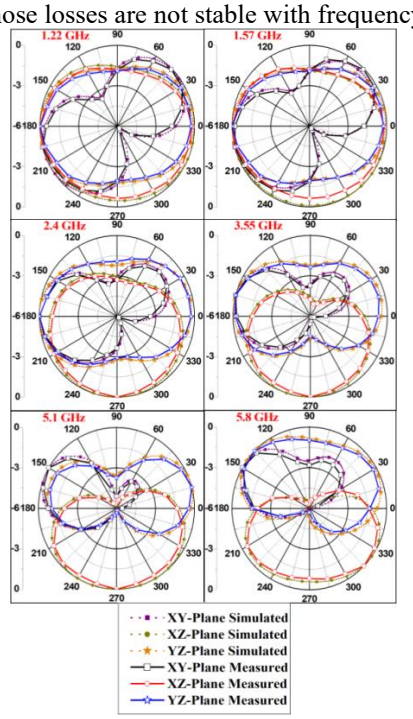


Fig. 9. Normalized radiation patterns of the proposed antenna *Ant-6*

E. Diversity Parameters

The diversity parameters are investigated to confirm the suitability of the proposed antenna as a MIMO antenna. The formulas are referred to as given in [27]. The envelope corrélation coefficient (ECC) of the proposed antenna is below 0.03 within the entire operating band, and the same characteristics are demonstrated in Fig 10 (a). Moreover, the diversity gain (DG) is calculated and found around the value of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:41:21 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Components, Packaging and Manufacturing Technology. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/TCPMT.2023.3304490

10 dB. In the continuation of the study of the diversity parameters, the proposed antenna's total active reflection coefficient (TARC) is also investigated which is lying below -20 dB for the working band. ECC and TARC are calculated using Eq (6) and Eq (9) of [30], respectively. The measurement of these parameters is done in anechoic chamber through vector network analyzer (VNA). Calculation of ECC is based on far-field parameters while TARC is depending on the S-parameters. The variation of DG and TARC concerning operating frequency is shown in Fig 10 (b), and all the values are found suitable for the MIMO performance.

Fig 10 (c) shows the simulated channel capacity loss (CCL) of *Ant-6* in terms of the number of bits per second per hertz (bits/s/Hz), and it is observed that CCL is of minimal value, which means the proposed antenna will support higher channel capacity. The calculation of CCL is done as per the formulas given in references [30,32-33]. The efficiency of the proposed antenna is studied in terms of radiation efficiency and total efficiency, shown in Fig 10 (d). The proposed antenna's radiation efficiency varies between 79% to 96.5%, whereas its total efficiency is found in the range of 79% to 95% (simulated) within the operating frequency. Measured total efficiency is found in accordance with the simulated results, as shown in Fig 10 (d).

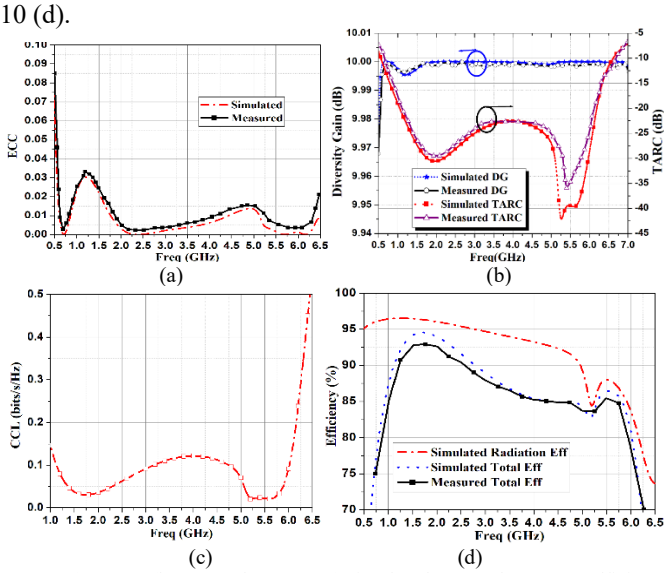


Fig. 10. (a) ECC, (b) DG and TARC, (c) simulated CCL, (d) antenna efficiency

A summary of the results of the proposed antenna (i.e. *Ant-6*) and its comparison with literature is comprised in TABLE I.

TABLE I
CHARACTERISTIC COMPARISON

Ref	Design Compacity	Operating range (GHz)	Efficiency (%)	Isolation (dB)	CP Band	Applied design method	L-Band support
[11]	$0.3\lambda_0 \times 0.7\lambda_0$	2.1–2.9	85–95	> 20	No	1.Orthogonal antenna placement 2.Decoupling structure	No
[14]	$0.3\lambda_0 \times 0.3\lambda_0$	1–6	78	> 20	No	Slotted Antenna with E &U–shaped slots	Yes
[15]	$0.2\lambda_0 \times 0.4\lambda_0$	3.34–3.87	92	> 15	No	1.Dipole Antenna 2. PIN Diodes	No
[16]	$0.7\lambda_0 \times 0.8\lambda_0$	1.8–2.9	60	> 18	No	1.Electric Dipole 2.Square Magnetic slot	No
[17]	$0.5\lambda_0 \times 0.5\lambda_0$	5.5–12.5	95	> 12	No	1.Bow-tie antenna 2.Partial semi ground plane	No
[28]	$0.3\lambda_0 \times 0.2\lambda_0$	2.3–2.5 & 5–5.19	82–90 & 5–66	> 20	No	1.Orthogonal antenna 2.Decoupling structure	No
[29]	$0.2\lambda_0 \times 0.3\lambda_0$	2–3	74	> 44	No	H-shaped DGS	No
Ant-6	$0.1\lambda_0 \times 0.1\lambda_0$	**0.97-6.19**	**79-95**	**> 23**	**integrated**	**Slots, stubs, defected ground, and parasitic elements**	**Yes**

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:41:21 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Components, Packaging and Manufacturing Technology. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/TCPMT.2023.3304490

IV. Conclusion

A wideband two-port MIMO antenna is designed, that covers the whole sub-6 GHz band ranging from 0.97 - 6.19 GHz with an integrated circularly polarized wideband centred at 3.5 GHz. A built-in isolator is integrated within the proposed structure, which provides more than 23dB isolation between ports. The proposed structure is compact and of $0.113\lambda_0 \times 0.129\lambda_0$ size. An improved CCL of about 0.12 bits/s/Hz is achieved along with other satisfactory diversity parameters, i.e. ECC, TARC, and DG. The omnidirectional pattern is achieved at the L1 (1.57 GHz) and L2 (1.22 GHz) bands of the Global Positioning System (GPS). Except for the L1 and L2 bands, the proposed antenna offers wide beamwidth for the remaining bands. Experimental results validate the simulated results.

REFERENCES

[1] FCC, "America's 5G Future," https://www.fcc.gov/5G.

[2] FCC "FCC Opens 6 GHz Band to Wi-Fi and Other Unlicensed Uses," Apr 23, 2020, https://www.fcc.gov/document/fcc-opens-6-ghz-band-wi-fi-and-other-unlicensed-uses.

[3] Federal Communications Commission's Fact Sheet on "Facilitating 5G in the 3.45-3.55 GHz Band," 09th Sept 2020, https://docs.fcc.gov/public/attachments/DOC-366780A1.pdf.

[4] FCC, The FCC's 5G Fast Plan, https://www.fcc.gov/5G (last visited Sept. 7, 2020).

[5] GSMA WRC Series - 3.5 GHz in the 5G Era: Preparing for New Services in 3.3-4.2 GHz, April 2021, https://www.gsma.com/spectrum/wp-content/uploads/2021/04/3.5-GHz-for-5G.pdf.

[6] Prakash Moorut, "5G at 3.5 GHz: The Global Opportunity," 29 Oct 2019, https://www.nokia.com/blog/5g-35-ghz-global-opportunity/.

[7] M.K. Khandelwal, "Metamaterial based circularly polarized four-port MIMO diversity antenna embedded with slow-wave structure for miniaturization and suppression of mutual coupling," *AEU - International Journal of Electronics and Communications*, vol. 121, no. 153241, pp. 1-10, July 2020.

[8] R. Chandel, A. K. Gautam and K. Rambabu, "Design and Packaging of an Eye-Shaped Multiple-Input–Multiple-Output Antenna with High Isolation for Wireless UWB Applications," in *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 8, no. 4, pp. 635-642, April 2018.

[9] M. Agarwal, J.K. Dhanoa, and M.K. Khandelwal, "Ultrawide band two-port MIMO diversity antenna with triple notch bands, stable gain and suppressed mutual coupling*," AEU - International Journal of Electronics and Communications (Elsevier)*, Vol 120, 2020.

[10] Wang, Enhao, Weimin Wang, Xiaohua Tan, Yongle Wu, Jinchun Gao, and Yuanan Liu, "A UWB MIMO slot antenna using defected ground structures for high isolation," *International Journal of RF and Microwave Computer-Aided Engineering* 30, no. 5 (2020): e22155.

[11] Ali, Wael AE, and Ahmed A. Ibrahim, "A compact double-sided MIMO antenna with an improved isolation for UWB applications," *AEU-International Journal of Electronics and Communications* 82 (2017): 7-13.

[12] M. Agarwal, J.K. Dhanoa, and M.K. Khandelwal, "Two-Port Hexagon Shaped MIMO Microstrip Antenna for UWB Applications Integrated with Double Stop Bands for WiMax and WLAN," *AEU - International Journal of Electronics and Communications (Elsevier),* Vol 138, pp. 153885, July 2021.

[13] Saurabh, Arun Kumar, and Manoj Kumar Meshram. "Compact sub-6 GHz 5G-multiple-input-multiple-output antenna system with enhanced isolation," *International Journal of RF and Microwave Computer-Aided Engineering* 30, no. 8 (2020): e22246.

[14] Desai, Arpan, Riki Patel, Trushit Upadhyaya, Hemani Kaushal, and Vigneswaran Dhasarathan, "Multiband inverted E and U shaped compact antenna for Digital broadcasting, wireless, and sub 6 GHz 5G applications," *AEU-International Journal of Electronics and Communications* 123 (2020): 153296.

[15] G. Jin, C. Deng, Y. Xu, J. Yang and S. Liao, "Differential Frequency-Reconfigurable Antenna Based on Dipoles for Sub-6 GHz 5G and WLAN Applications," in *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 3, pp. 472-476, March 2020.

[16] M. Sonkki, E. Antonino-Daviu, M. Cabedo-Fabres, M. Ferrando-Bataller and E. T. Salonen, "Improved Planar Wideband Antenna Element and Its Usage in a Mobile MIMO System," in *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 826-829, 2012.

[17] A. A. Eldek, A. Z. Elsherbeni and C. E. Smith, "Wide-band modified printed bow-tie antenna with single and dual polarization for C - and X-band applications," in *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 9, pp. 3067-3072, Sept. 2005.

[18]M. Alibakhshikenari et al., "A Comprehensive Survey on "Various Decoupling Mechanisms With Focus on Metamaterial and Metasurface Principles Applicable to SAR and MIMO Antenna Systems"," in IEEE Access, vol. 8, pp. 192965-193004, 2020, doi: 10.1109/ACCESS.2020.3032826.

[19]M. Alibakhshikenari, Bal S. Virdee, Ernesto Limiti, "Study on isolation and radiation behaviours of a 34×34 array-antennas based on SIW and metasurface properties for applications in terahertz band over 125–300 GHz," Optik,Vol 206, 163222, 2020.

[20]M. Alibakhshikenari, et. al., "Isolation enhancement of densely packed array antennas with periodic MTM-photonic bandgap for SAR and MIMO systems," IET Microw. Antennas Propag., vol. 14, pp. 183-188, 2020, doi: https://doi.org/10.1049/iet-map.2019.0362.

[21]M. Alibakhshikenari, et. al., "Surface Wave Reduction in Antenna Arrays Using Metasurface Inclusion for MIMO and SAR Systems," Radio Science, vol. 54, pp. 1067–1075, 2019, doi: https://doi.org/10.1029/2019RS006871.

[22]M. Alibakhshikenari, M. Khalily, B. S. Virdee, C. H. See, R. A. Abd-Alhameed and E. Limiti, "Mutual-Coupling Isolation Using Embedded Metamaterial EM Bandgap Decoupling Slab for Densely Packed Array Antennas," in IEEE Access, vol. 7, pp. 51827-51840, 2019, doi: 10.1109/ACCESS.2019.2909950.

[23]M. Alibakhshikenari, M. Khalily, B. S. Virdee, C. H. See, R. A. Abd-Alhameed and E. Limiti, "Mutual Coupling Suppression Between Two Closely Placed Microstrip Patches Using EM-Bandgap Metamaterial Fractal Loading," in IEEE Access, vol. 7, pp. 23606-23614, 2019, doi: 10.1109/ACCESS.2019.2899326.

[24]M. Alibakhshikenari, et al., "Interaction between closely packed Array antenna elements using meta-surface for applications such as MIMO systems and synthetic aperture radars," Radio Science, vol. 53, pp. 1368–1381, 2018. https://doi.org/10.1029/2018RS006533.

[25]M. Alibakhshikenari, et al., "Antenna Mutual Coupling Suppression Over Wideband Using Embedded Periphery Slot for Antenna Arrays," Electronics, vol. 7, 198, 2018, https://doi.org/10.3390/electronics7090198.

[26]M. Alibakhshikenari, B.S. Virdee, C.H. See, R. Abd-Alhameed, A. Hussein Ali, F. Falcone, and E. Limiti, "Study on isolation improvement between closely-packed patch antenna arrays based on fractal metamaterial electromagnetic bandgap structures," IET Microw. Antennas Propag., vol. 12, pp. 2241-2247, 2018, doi: https://doi.org/10.1049/iet-map.2018.5103.

[27]M. Alibakhshikenari, B.S. Virdee, P.K. Shukla, C.H. See, A. Raed. F. Falcone, and E. Limiti, "Meta-Surface Wall Suppression of Mutual Coupling Between Microstrip Patch Antenna Arrays for THz -Band Applications," Progress In Electromagnetics Research Letters, Vol. 75, 105-111, 2018, doi:10.2528/PIERL18021908.

[28]D. Sipal, M. P. Abegaonkar and S. K. Koul, "Highly Isolated Compact Planar Dual-Band Antenna With Polarization/Pattern Diversity Characteristics for MIMO Terminals," in *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 4, pp. 762-766, April 2019.

[29]Acharjee, Juin, Kaushik Mandal, and Sujit Kumar Mandal, "Reduction of mutual coupling and cross-polarization of a MIMO/diversity antenna using a string of H-shaped DGS," *AEU-International Journal of Electronics and Communications*, vol 97, pp 110-119, 2018.

[30]M Agarwal, JK Dhanoa, and MK Khandelwal, "Construction of compact two-port throne-shaped antenna with dual stopbands and high isolation for Ultrawideband Multiple-Input Multiple-Output wireless communication systems," *Int J Commun Syst.*, 2022; 35 (11): e5195.

[31]C. A. Balanis, "Antenna.Theory.Analysis.and.Design,"Wiley, 2016.

[32]Sharawi MS. Printed multi-band MIMO antenna systems and their performance metrics [wireless corner]. IEEE Antennas and Propagation Magazine. 2013;55(5):218-232. doi:10.1109/MAP.2013.6735522.

[33]Karaboikis MP, Papamichael VC, Tsachtsiris GF, Soras CF, Makios VT. Integrating compact printed antennas onto small diversity/MIMO terminals. IEEE Transactions on Antennas and Propagation. 2008;56(7):2067-2078. doi:10.1109/TAP.2008.924677.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:41:21 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Components, Packaging and Manufacturing Technology. This is the author's version which has not been fully edited content may change prior to final publication. Citation information: DOI 10.1109/TCPMT.2023.3304490

Monika Kaushik is pursuing Ph.D. Degree in Wide Band MIMO Antennas from IGDTUW, Delhi, India. She is also working as Assistant Professor in the dept. of Electronics & Communication Engineering of Bhagwan Parshuram Institute of Technology, Delhi. Her current research is focused on microstrip antennas, and 5G MIMO antennas.

Jasdeep K. Dhanoa (Senior Member, IEEE) did her Ph.D. in Electronics Engineering, from GGS Indraprastha University Delhi India, M. Tech in Electronics & Communication Engineering from Delhi College of Engineering and B.E. in Electronics Engineering.

She is a professor in the dept of Electronics & Communication Engineering of IGDTUW, New Delhi India. She also holds the post of Dean (Academic Affairs) of the University. She joined IGDTUW in 2014. She has a teaching experience of over 18 years at GTBIT, (affiliated GGSIPU) and Netaji Subhas Institute of Technology.

Prof. Dhanoa is an active member of ISTE, IETE, IEEE and WE00123. She is Vice Chair of WE00123 - Delhi Section Affinity Group, WIE-2015 and Chair Educational Activities of IEEE Delhi Section, 2015.

Mukesh K. Khandelwal (Senior Member, IEEE) received his Ph.D Degree in Electronics Engineering from Indian Institute of Technology ISM, Dhanbad, India in 2015.

He is working as an Assistant Professor in the School of Engineering of Jawaharlal Nehru University, Delhi, India. He has served as Deputy Director R & D of the University from Sept-2020 to Sept-2022. Moreover, he has also worked as Assistant Professor at GGSIPU and UPTU. His current research interests include microstrip antennas, metamaterials, 5-G antennas, UWB antennas, rectennas, reconfigurable antennas, multiband antennas, circularly polarized antennas, antenna array, defected ground structure and microwave components.

Dr Khandelwal has been received SERB International Research Experience (SIRE) fellowship awarded by Science and Engineering Research Board (SERB), a statutory body under the Department of Science and Technology, Government of India. He is serving as a reviewer in many reputed journals. He has published more than 40 SCI journals' publications and several international conference papers.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:41:21 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Performance Enhancement of a PMSM Drive using ANFIS- HCC for Electric Vehicles

Brijendra Sangar (brijendra_2k21phdee30@dtu.ac.in)
Delhi Technological University
Madhusudan Singh
Delhi Technological University
Mini Sreejeth
Delhi Technological University

Research Article

Keywords: Permanent Magnet Synchronous Motors (PMSM), Hysteresis Current Control (HCC), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Field Oriented Control (FOC), Electric Vehicle (EV).

Posted Date: August 24th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3281264/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Performance Enhancement of a PMSM Drive using ANFIS-HCC for Electric Vehicles

Brijendra Sangar
Electrical Engineering Department
Delhi Technological University
New Delhi, India
brijendra_2k21phdee30@dtu.ac.in

Madhusudan Singh
Electrical Engineering Department
Delhi Technological University
New Delhi, India
madhusudansingh@dce.ac.in

Mini Sreejeth
Electrical Engineering Department
Delhi Technological University
New Delhi, India
minisreejeth@dce.ac.in

***Abstract*- Modern motor control algorithms have greatly enhanced the efficient operation of Permanent Magnet Synchronous Motors (PMSM). One such creative development is the use of Adaptive Neuro-Fuzzy Inference Systems (ANFIS) control for PMSM motors. This combines the advantages of neural networks and fuzzy logic for the control of PMSM drives, resulting in enhanced efficiency, better dynamic performance, and accurate control of the motor. This paper explores the benefits and application of ANFIS control in PMSM drives, highlighting its potential to power the upcoming generation of electric drive systems in the e-vehicle (EV) industries. ANFIS control is included into the drive system to implement Field Oriented Control (FOC) of PMSM. The performance of the PMSM drive is evaluated using stator voltage and torque equations under different speeds and load torque conditions. A comparative analysis of the performance parameters is presented for the conventional Proportional Integral (PI) control coupled with Hysteresis Current Control (HCC) and ANFIS speed controller integrated with HCC. The nonlinearity of PMSM makes it challenging to employ the standard PI gain settings, which results in undesired overshoot. The designed and developed ANFIS-HCC controller is found to overcome this problem and it provides consistent enhanced performance characteristics. Simulation studies and experimental validation of the proposed novel controller design bears testimony to its possible adoption in new-generation EVs owing to enhanced performance and improved dynamic characteristics.**

***Keywords*— Permanent Magnet Synchronous Motors (PMSM), Hysteresis Current Control (HCC), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Field Oriented Control (FOC), Electric Vehicle (EV).**

1. INTRODUCTION

Replacement of conventional internal combustion (IC) engine vehicles with EVs can drastically reduce automotive pollution. The motors used for EV drive application need to have flexible torque and speed characteristics, to cover the entire vehicle operating range [1]. They also need to be very efficient to get an acceptable driving range per charge cycle, in comparison with conventional IC engine vehicles. To cater to the needs of urban driving with an acceptable driving range and lower operating costs, it is crucial to improve motor efficiency under low-torque operating conditions [2].

A typical EV propulsion system is divided into electrical and mechanical components. Battery packs are used in the electrical system to store and supply power. Low voltage is converted to high voltage levels using the dc-dc boost converter. Direct current (dc) is transformed into alternating current (ac) by the voltage source inverter (VSI), which powers the machines. Further, a breaker connecting the battery packs and the converter is used for protection, and a bus capacitor in parallel with the inverter is used for high-frequency power surge absorption. A gearbox, usually connected to the PMSM shaft, serves as a differential, clutch, and speed governor. In the propulsion system, the protection mode is activated instantly in case of emergencies like vehicle crashes, etc.

PMSM drives are widely utilized in applications that require better torque and speed control, high power density, high torque to inertia ratio, and efficient drive systems. Unlike DC machines` PMSMs do not have brushes, which makes them more reliable. The recent past has witnessed increasing use of PMSM drives in applications traditionally dominated by dc machines, such as robotics, power, gas, oil, and transportation sectors. Induction motors, PMSMs, and switched reluctance motors have been traditionally used in EV propulsion systems [3]. However, PMSMs remain the popular choice due to their higher power density, ability to run at constant power for the majority of its operational range, and least acoustic vibration and noise. Furthermore, its high efficiency can be maintained across a wide range of speed-torque characteristics [4][5].

The development of effective control methods has resulted in lower torque ripples in PMSM drives. The most widely employed control in PMSM drives is the FOC technique [6][7][8]. In FOC, torque and flux are decoupled through field orientations and regulating the phase currents in the stator [9]. General linear control methods, however, are unable to satisfy the demands of high-performance applications of PMSM control systems due to external disturbances and inevitable errors in the model. A plethora of non-linear control techniques have been reported in the literature for performance enhancement, which includes sliding mode control [10],[11] neural network control [12][13], adaptive control [14][15], linearization control [16] and fuzzy control [17].

These control methods minimize losses and offer a reliable response, when parameter fluctuations occur, without depending on steady-state processes. The conventional PI controller for a PMSM drive has a limited dynamic characteristic range. Fine-tuning the PI controller also becomes difficult due to the linearized approximation of function at any given point. The ANFIS controller, on the other hand, may select the entire system and adjust to it. When used in low-power applications, the

conventional PI controller often responds quickly, however PI tuning cannot be used to increase drive performance in high-power applications. Replacing the conventional PI controller with an ANFIS controller in FOC-based PMSM drives overcomes this drawback thus producing better dynamic performance [18].

The combination of fuzzy inference with neural networks provides a versatile technique for designers to effectively create and operate electric drives due to its non-linearity handling capacity [19]. The ability of artificial neural networks (ANN) to reliably predict complicated correlations between a collection of inputs and outputs [20], which can be extremely nonlinear, unexpected, and unpredictable, has led to their increasing use in a variety of fields

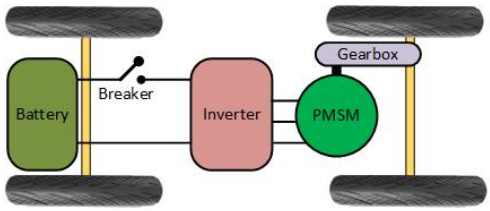


Fig. 1. Schematic of the PMSM drive for Electric vehicle control

This paper presents the design and implementation of an ANFIS-HCC speed controller for minimizing torque ripples in FOC-based PMSM drives. The ANFIS-HCC controller has been developed through simulation studies using MATLAB / Simulink and the laboratory prototype is implemented using dSPACE ds1104 controller. The schematic diagram of the system is shown in Fig. 1. The electric drive includes the battery, VSI and PMSM motor. The inverter is controlled by the ANFIS-HCC controller. The developed controller generates the reference speed and eliminates ripples by reducing system non-linearities. Experimental validation of the developed ANFIS-HCC controller is done and the comparative analysis of its performance vis a vis the conventional PI-HCC controller for FOC of PMSM drive is presented.

The novelty of this paper can be summarized as: -

a. The ANFIS-HCC for FOC of PMSM drive minimizes the torque and current ripples, resulting in the enhanced static and dynamic performance of the drive under diverse operating conditions of an EV.

b. The performance of the developed controller is verified using the EV industry-standard Inspection and Maintenance Driving Cycle (IM240).

This paper is organized into six sections, with an Introduction followed by Section II, which presents the mathematical models of EV and PMSM, Section III describes the ANFIS and HCC design. Section IV covers the Simulation Studies and Section V presents the experimental results and discussions. Section VI concludes the paper.

2. Mathematical Modeling

A. Mathematical Modelling of PMSM- based EV

The resistance forces acting on a vehicle in motion on a gradient are shown in Fig. 2. This is modeled mathematically as given in (1) and (2) and vehicle parameters as given in Table I.

$$F_T = F_R + F_W + F_G + F_A \quad (1)$$

$$F_T = K_R \mathrm{W}\cos(\delta) + K_W A V^2 + \mathrm{W}\sin(\delta) + W\frac{a}{g} \quad (2)$$

Where, F_T is the total tractive force of the vehicle, F_w the wind resistance, F_G the grade resistance, F_R is the rolling resistance due to road and tire interaction, and F_A is the accelerating force generated from the vehicle mass. K_R is the Rolling resistance coefficient, K_W is the wind resistance coefficient, A is the Vehicle frontal area, V is the velocity of the vehicle, W is the weight of the vehicle, δ is the gradient of the road, a is the acceleration of vehicle, g is the acceleration due to gravity.

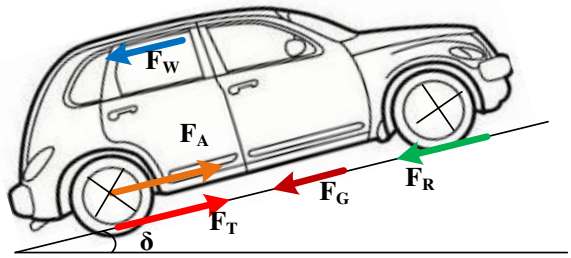


Fig. 2. Resistive forces acting on the vehicle

The motor torque and speed for various operating conditions are estimated using (3) and (4)

$$\omega_r = \mathrm{V}\frac{60 * r_f}{2 * pi * r_r} \quad (3)$$

$$\tau_L = \frac{r_r}{r_f} F_T \quad (4)$$

Where ω_r is the motor speed in rpm; V the vehicle speed in m/sec ; r_f is the gear ratio and r_r is the rolling radius of the wheel in meter.

Unlike conventional testing of electric motor drives meant for industrial applications, which is usually tested under constant or step or ramp changes in speed/torque conditions, the testing of motor drives for EV applications is tested using predetermined drive cycles, which represent the actual driving conditions depending on traffic, road condition, road gradient, etc. [21] [22] [23]. Therefore, a powertrain model should be constructed to test the drive depending on the required driving cycle. The Inspection and Maintenance Driving Schedule (IM240) is used to test the performance of the developed controller for PMSM used in the power train of EVs.

Table I. Vehicle parameters.

Vehicle weight (p.u.), W	1.0
Vehicle frontal area, A	2.35m^2
Wheel rolling radius, r_r	0.3048m
Gear ratio, r_f	4.5
Rolling resistance coefficient, K_r	0.015
Wind resistance coefficient, K_w	0.3

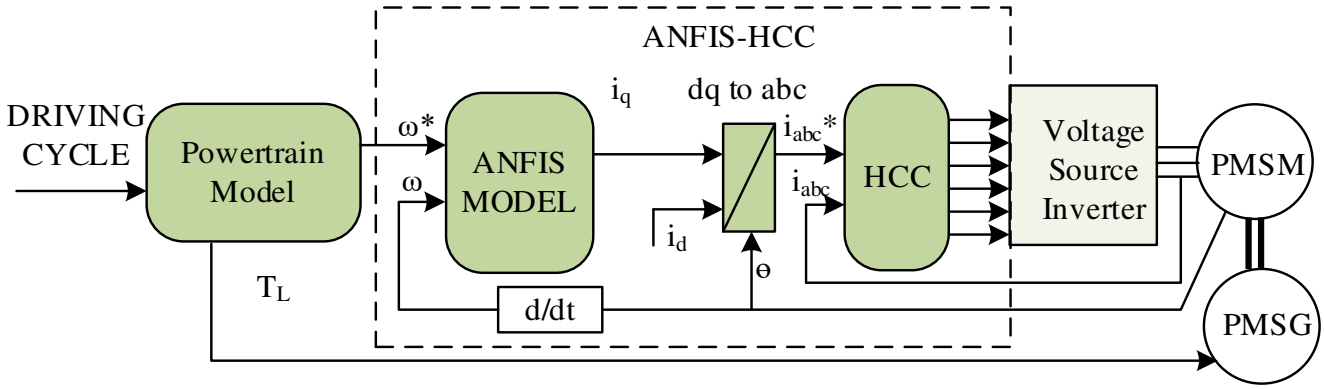


Fig. 3. Block Diagram of the Proposed ANFIS-HCC.

A. Modelling of PMSM drive

The vector control technique is used to decouple the stator current to create current components of direct and quadrature axes. The voltage equations of Surface Mounted PMSM (SPMSM) in the rotor reference frame are given in the (5), (6), and (7). Where v_q and V_d, respectively, are the stator quadrature and direct-axis voltage components; i_q and i_d , respectively, are the stator quadrature and direct-axes current components; L_s is the stator inductance/phase; R is the stator resistance/phase; ψ_f is the rotor flux linkage and ω_s is the rotor speed in electrical rad/s, v_a, v_b, and v_c sre the phase voltages.

$$\begin{bmatrix} v_q \\ v_d \\ v_0 \end{bmatrix} = \frac{2}{3}\begin{bmatrix} \cos\theta & \cos(\theta-120) & \cos(\theta+120) \\ \sin\theta & \sin(\theta-120) & \sin(\theta+120) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} v_a \\ v_b \\ v_c \end{bmatrix} \quad (5)$$

$$v_q = R \times i_q + L_s\frac{di_q}{dt} + L_s\omega_s i_d + \omega_s \times \psi_f \quad (6)$$

$$v_d = R \times i_d + L_s\frac{di_d}{dt} - L_s\omega_s i_q \quad (7)$$

The direct and quadrature axes stator currents are expressed as given in (8) and (9)

$$i_d = \frac{1}{L_s}\int [V_d - R \times i_d + \omega_s \times \psi_f]dt \quad (8)$$

$$i_q = \frac{1}{L_s}\int [V_q - R \times i_q - \omega_s \times \psi_f]dt \quad (9)$$

The developed torque and speed are given in (10) and (11) respectively. Where T_L is the load torque; P_n, is the number of poles; B, is the motor friction coefficient, T_e, is the electromagnetic torque; and ω_r is the rotor's mechanical speed.

$$T_e = 1.5 \times P_n \times \psi_f \times i_q \quad (10)$$

$$\omega_s = P_n \times \omega_r \quad (11)$$

3. ANFIS-HCC Controller for EV

The ANFIS model controls the outer loop of the FOC-based PMSM drive, whereas the HCC controls the inner current loop, as shown in Fig. 3. The control design strategy is elucidated in this section.

A) Adaptive Neuro-Fuzzy Inference System

ANFIS is an ANN that works with the Takagi-Sugeno fuzzy inference system [24],[25]. The technique builds the fuzzy inference system from an input-output pair and an observation data set that has been provided. In the learning phase, the backpropagation technique is used to modify the membership function's parameters in the designed ANFIS architecture.

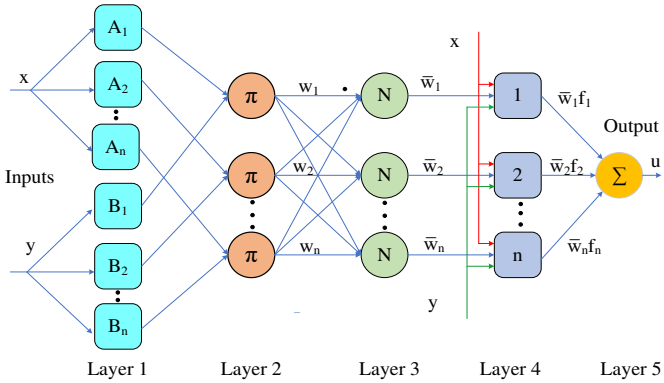


Fig. 4. Block Diagram of ANFIS

As shown in Fig. 4, the ANFIS structure is divided into five layers. The nodes in Layer 1 are adaptive nodes. The fuzzy membership grade of the inputs is given by (12) and (13)

$$O_{1Ai} = \mu_{Ai}(x), i = 1,2,3 \ldots n \quad (12)$$

$$O_{1Bi} = \mu_{Bi}(y), i = 1,2,3 \ldots n \quad (13)$$

Where x and y are the inputs to node i; Ai and Bi are the linguistic labels (high, low, etc.) connected to this node function μ. Each fuzzy membership function is adaptable by A_i and B_i , which depends on the least error. A triangular membership function is chosen as the shape with the least amount of error. It is specified by three variables: a and c for the feet and b for the curve tip. The triangular function is expressed as given in (14)

$$\text{triangular}\,(x,a,b,c) = max\left(min\left(\frac{x-a}{b-a}, 1, \frac{c-x}{c-b}\right), 0\right) \quad (14)$$

Each node with a label of π corresponds to the firing strength of

a fuzzy rule in Layer 2 (inference layer). The output of the Inference Layer is given in (15)

$$O_{2,i} = w_i = A_i(x) \times B_i(y) \quad , i = 1,2,3 \ldots n \tag{15}$$

The third layer, a normalization layer, normalizes the estimated firing strengths from the previous layer using (16).

$$O_{3,i} = \overline{w}_i = \frac{w_i}{\sum_i (w_i)} \quad , i = 1,2,3 \ldots . n \tag{16}$$

The nodes in Layer 4, are adaptive. The output of each node in this layer is the product of the normalized firing strength and a first order polynomial (for a first order Sugeno model). The output of this layer is given by (17).

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i), i = 1,2,3 \ldots . n \tag{17}$$

where f_i is the control signal and $(p_i, \ q_i, \ r_i)$ is the consequence parameter set. The output layer calculates the sum of all incoming signals and de-fuzzifies the subsequent set of rules using (18)

$$O_{5,i} = \sum_i (\overline{w}_i f_i) = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad i = 1,2,3 \ldots . . n \tag{18}$$

B) Training of ANFIS

ANFIS controller must be trained and optimized before it can be used. To generate the data for training of the developed ANFIS Controller, the PMSM is driven using a conventional PI-HCC controller. Data corresponding to 50,001 pairs of input-output, captured by the control desk tool in dSPACE, were used to train the ANFIS-HCC using the triangular membership function (MF). The Flowchart for Training of the proposed ANFIS-HCC is shown in Fig. 5. Hybrid approach is used for minimization of error, wherein backpropagation is combined with least-squares techniques for best accuracy. In this, the error tolerance is set at zero, and 50 epochs were chosen for optimal training accuracy and time.

C) Hysteresis Current Controller

High-performance vector-controlled drives require large current-control loop bandwidths to enable accurate current tracking, rapid transient response, and to compel the VSI to act as an equivalent current-generating amplifier within the current loop bandwidth. When achieving the torque or speed requirements, HCC is employed to provide an output current that is almost sinusoidal [26].

Using an HCC with a hysteresis band for VSI current control in the inner loop ensures a nearly sinusoidal waveform, which is achieved by comparing the actual stator current, (i_{abc}) to the reference motor current, (i^*_{abc}). Due to the reference current being within a predetermined bandwidth, this also guarantees a low harmonic content in the stator current. The error in current determines the activation of the appropriate switches of the VSI depending on its value. The switching frequency of the controller is band limited. The ANFIS controller creates the quadrature axis stator current, i_q, in the outer loop based on the difference between the reference speed $(\omega_r{}^*)$ and the actual speed (ω_r). This along with the direct axis stator current, i_d, is converted to i^*_{abc} using park transformation. This reference three-phase current (i_{abc}*) is fed along with the measured current (i_{abc}) to the HCC to generate the firing pulse for the VSI to drive the PMSM.

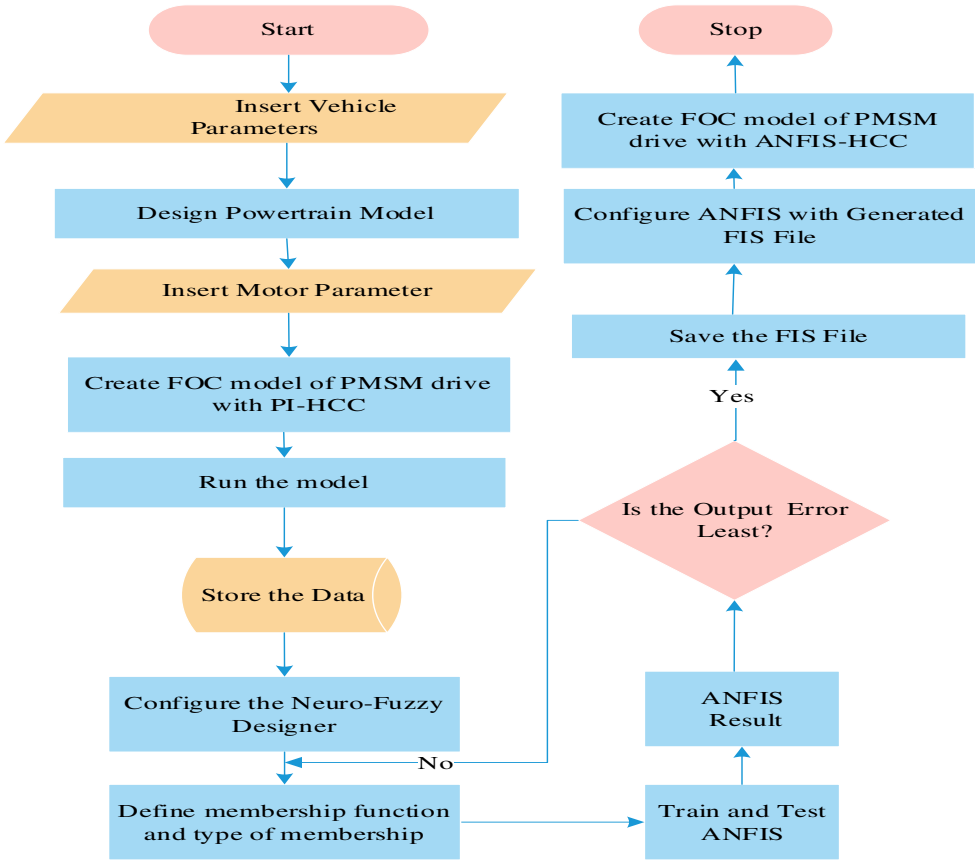


Fig. 5. Flowchart for Training of proposed ANFIS-HCC

4. SIMULATION STUDIES AND DISCUSSIONS

The proposed ANFIS-HCC controller is developed through simulation studies using MATLAB/Simulink with a sampling time of 5 µs for the motor with specifications as given in Table II. This section presents the comparison of the performance of the FOC based PMSM drive using conventional PI-HCC and the proposed ANFIS-HCC controller under different operating conditions.

The cascaded control system structure consists of two control loops. The outer loop is the speed control loop, and the inner loop is the current control loop. In the PI control structure, the outer-speed controller loop and the inner-current controller loop are based on PI and HCC controllers respectively. Whereas, the ANFIS-HCC control scheme is a combination of ANFIS in the speed control loop and HCC in the current control loop.

When a motor starts up, it requires a large amount of current to overcome inertia, but this current gradually decreases as the speed reaches a steady state value. The speed responses of the aforesaid two controllers are simulated and analyzed. The starting response of PMSM drive at rated-speed and no-load operation using PI-HCC and proposed ANFIS-HCC are presented in Figs. 6 and 7 respectively.

While starting with the reference speed set to 3000 rpm under no-load condition, some distortion and overshoot in electromagnetic torque (T_e) ,speed (ω_r), and stator current (i_a) are observed in PI-HCC based controller. However, the ANFIS-HCC based drive gives fast and smooth speed characteristics without overshoot. Fig. 8 shows the zoomed-in starting characteristics of the PMSM drive. In addition, the ANFIS-HCC controller has a shorter rising and settling time than the PI-HCC controller.

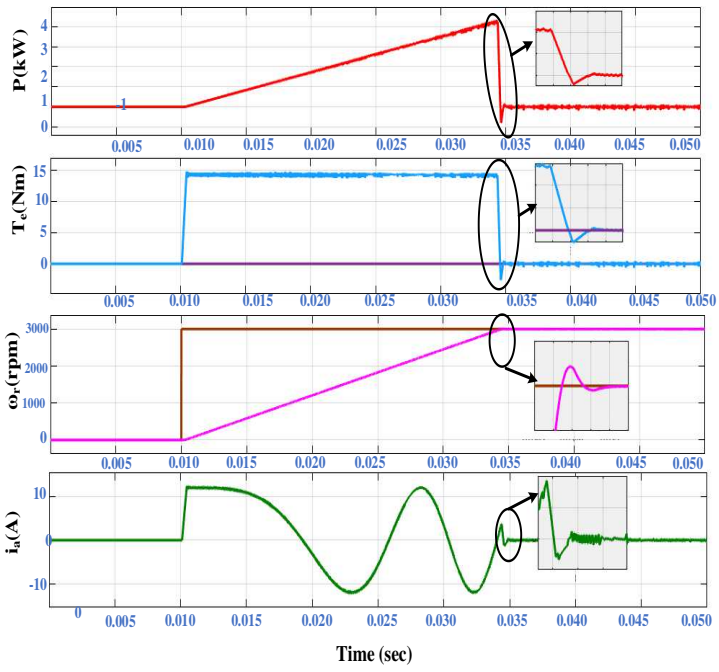


Fig. 6 Dynamic-Response of PI-HCC-based PMSM Drive under no-load rated speed (3000 rpm) Operation.

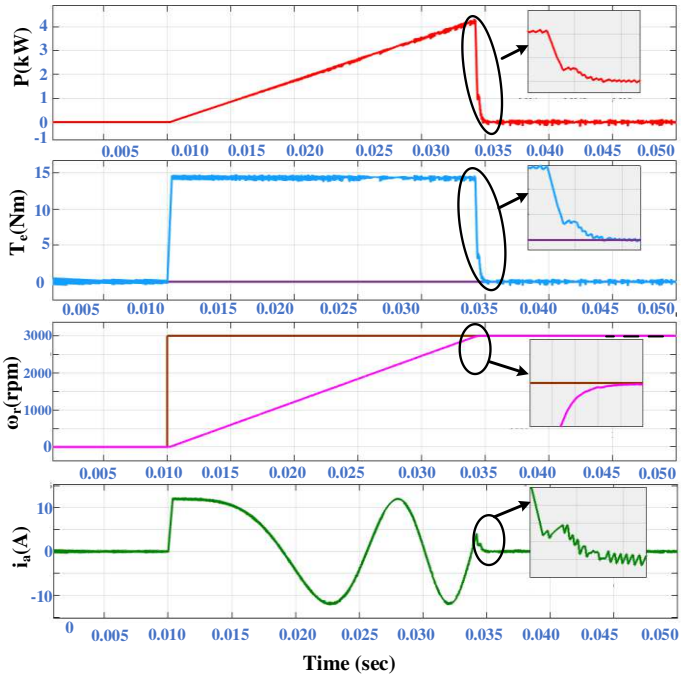


Fig. 7 Dynamic Response of ANFIS-HCC based PMSM Drive under no-load rated speed (3000 rpm) Operation

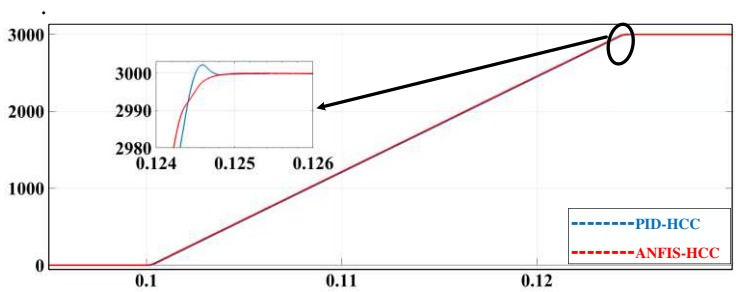


Fig. 8. Starting response of the PI-HCC and ANFIS-HCC based PMSM Drive under no-load-rated speed (3000 rpm) Operation

The proposed ANFIS-HCC for FOC of PMSM drive provides a significant reduction in current and torque ripples of PMSM drive in addition to fast and accurate speed tracking in comparison with the conventional PI-HCC-based controller.

5. EXPERIMENTAL RESULT AND ANALYSIS

A. *Experimental Setup*

A 5 HP, 380 V, 6 poles, and 3000 rpm SPMSM, with machine parameters as tabulated in Table II, is used for the experimental setup. The laboratory prototype, shown in Fig. 9, is powered by a bidirectional dc power supply. A dSPACE 1104 controller is programmed to control a 3-arm voltage source IGBT converter. MATLAB software is used to program the dSPACE 1104. Current sensors are used to detect alternating current. A permanent magnet synchronous generator (PMSG) coupled to the shaft of the SPMSM is used for electrical loading using resistor banks connected to the PMSG.

An incremental rotary encoder with a 1000-point base generates the feedback speed and position signal. An incremental encoder generates the A and B output signals using a quadrature

encoder. The pulses from the A and B outputs have square waveforms and a 90-degree phase difference because they are quadrature-encoded, even though the incremental encoder is working at a constant speed.

The phase difference between A and B signals will either be positive or negative depending on the direction the encoder is travelling at any particular time. Depending on the device's design, the phase difference for a rotary encoder is +90° for clockwise rotation and -90° for counterclockwise rotation. The frequency of the pulses on the A or B output is directly related to the encoder's velocity (rate of position change); greater frequencies indicate faster movement, while lower frequencies indicate slower movement. The frequency of a rotary encoder represents the pace of the encoder's shaft rotation, while the frequency of a linear encoder indicates the speed of linear traversal.

Table II. PMSM parameters

Rated Voltage	380 V
Rated speed, ω_r	3000rpm
Torque, T	11 Nm
No. of poles, P_n	6
Stator resistance, R	1.93 ohm
Rated Current, I	6.9A
Inductance, L_s	11.4mH
Flux linkage, ψ_f	0.265 wb
Voltage constant	144.7 Volts/Krpm
Torque constant, Kt	1.197 Nm/A
Inertia	0.0011 kg-m^2

B. *Starting Performance of PMSM drive*

The dynamic response of the laboratory prototype using conventional PI-HCC and the proposed ANFIS-HCC for FOC of PMSM Drive during startup under no load – rated speed (3000rpm) operation is shown in Figs. (10) and (11), respectively. The electromagnetic torque produced on reaching steady state speed enables the PMSM to achieve and maintain the reference rotor speed. It is observed that the FOC of PMSM using ANFIS-HSS takes 10% lesser rise time as compared to the conventional PI-HCC-based FOC of PMSM, as tabulated in Table III. In addition, the speed, torque, and stator currents exhibit some distortion in PI-HCC-based controller, unlike the ANFIS-HCC-based controller. The ANFIS-HCC controller is thus observed to offer a quick and smooth response.

Table III. Comparison of Starting Speed Response

Control Method	Rise Time (ms)	Settling Time (ms)
PI-HCC	293.4	482
ANFIS-HCC	264.7	428

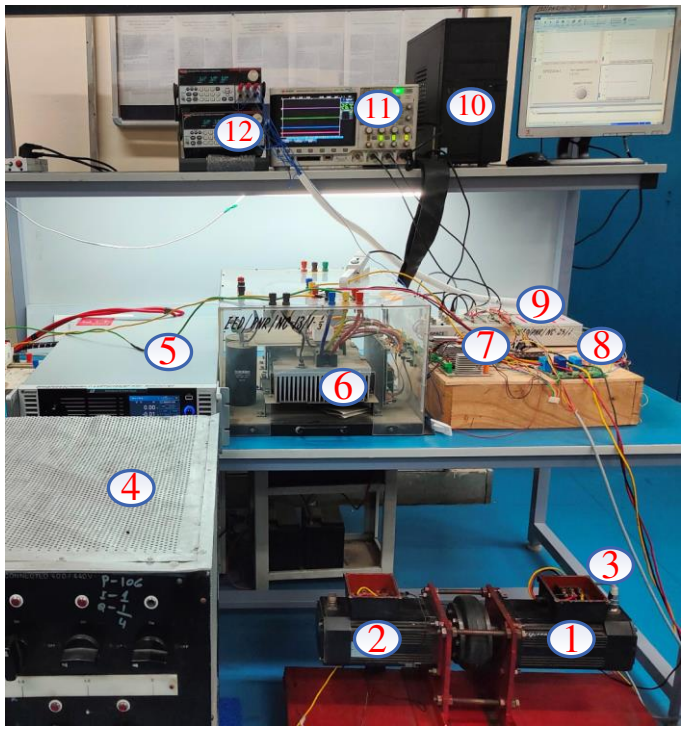


Fig. 9. Experimental setup for testing the ANFIS-HCC based PMSM drive: 1-PMSM, 2-PMSG, 3- Rotary Encoder, 4- Resistive load, 5- Bi-directional dc power supply, 6-Three-Phase Inverter, 7- Driver circuit, 8- Current Sensors, 9- dSPACE1104,10-PC, 11-4-Channel DSO, 12-dc power supply

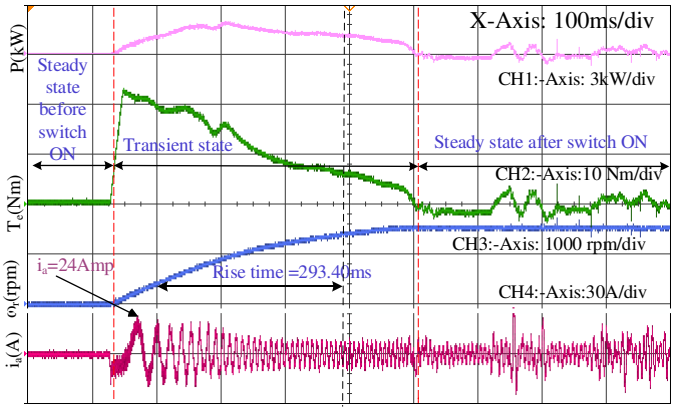


Fig. 10. Dynamic Response of Laboratory Prototype at Start-up under rated speed – no load operation with PI-HCC-based FOC

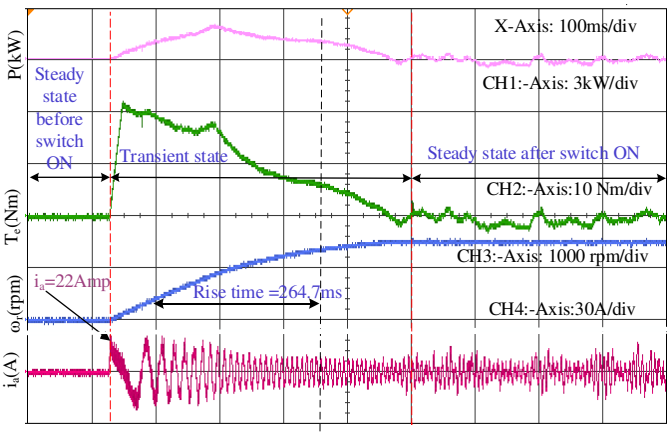


Fig. 11. Dynamic Response of Laboratory Prototype at Start-up under rated speed – no load operation with ANFIS-HCC based FOC.

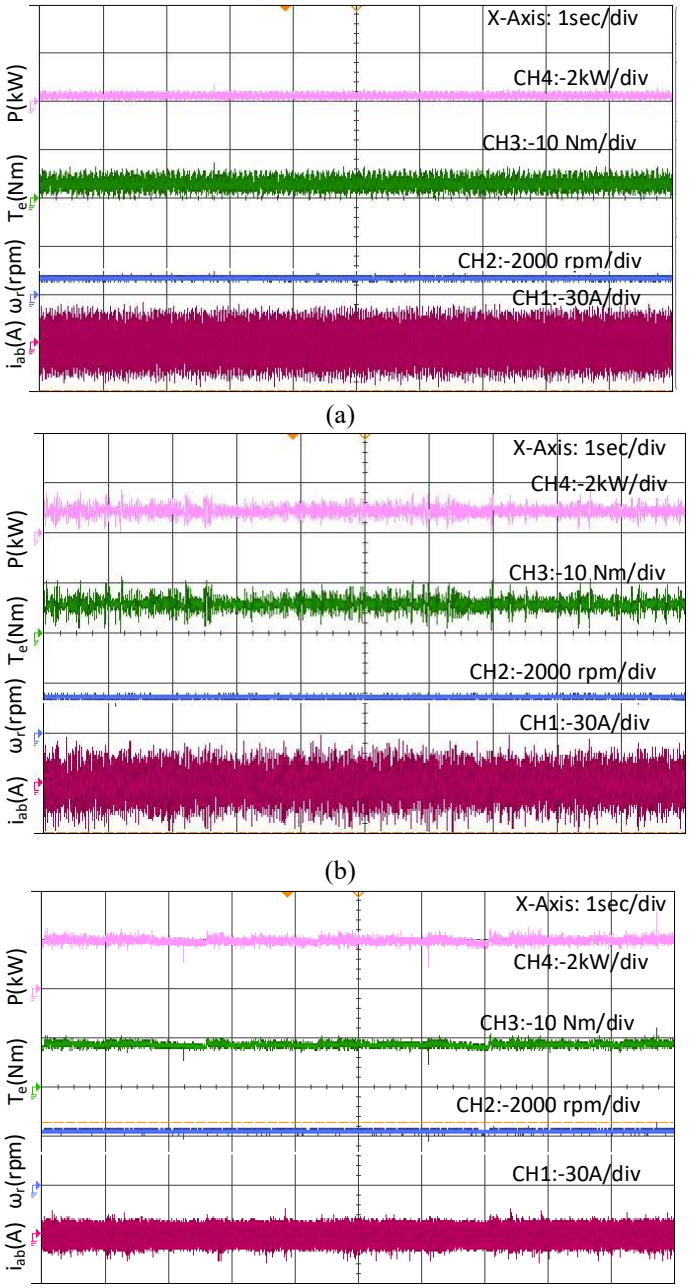


Fig. 12. Steady-state response of PI-HCC based PMSM drive at (a)750rpm-3Nm (b)1500rpm-6Nm(c)2250rpm-9Nm.

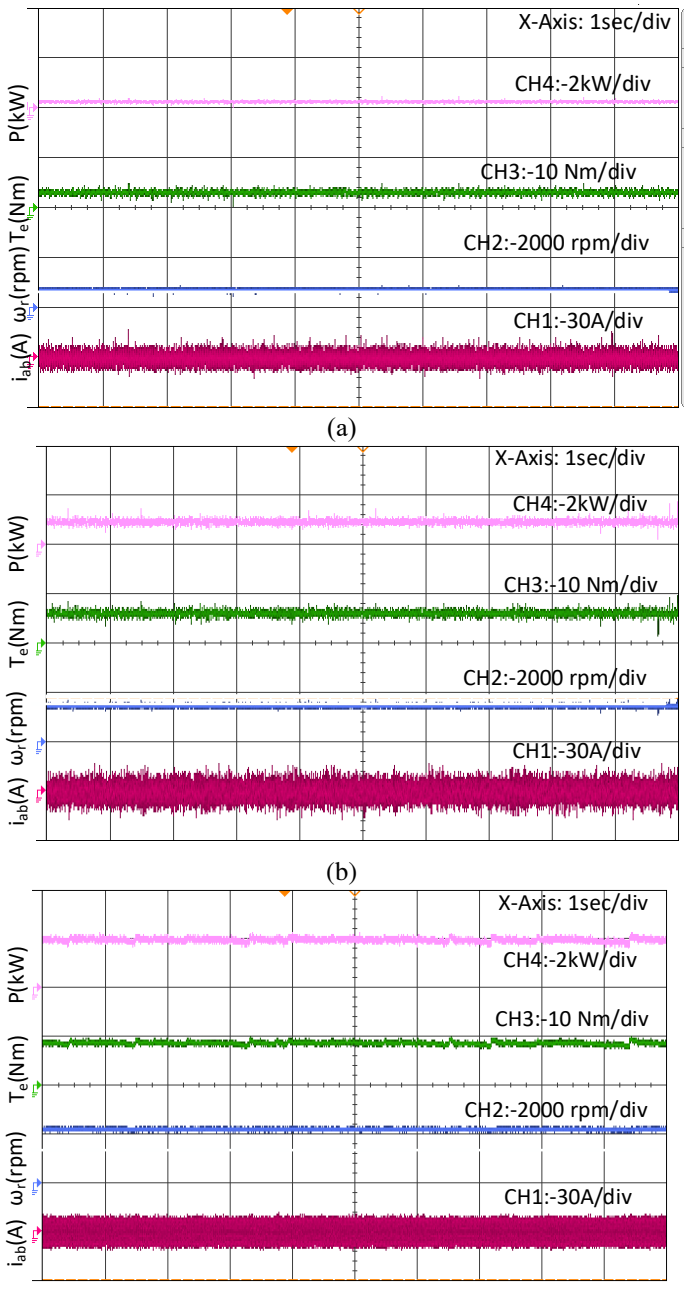


Fig 13. Steady-state response of ANFIS-HCC based PMSM drive at (a)750rpm-3Nm (b)1500rpm-6Nm(c)2250rpm-9Nm.

C. Steady-state performance of PMSM drive

To validate the effectiveness of the proposed ANFIS-HCC-based scheme for FOC of PMSM Drive, the Steady State performance is assessed under different speeds with load torque of 25%, 50%, and 75% of its rated load. The motor speed, electromagnetic torque, stator current, and input power for the PMSM drive using the conventional PI-HCC and proposed ANFIS-HCC are shown in Figs. 12 and 13, respectively. It is observed that the perturbations in power and current are less in the ANFIS-HCC-based FOC. Torque ripples in electromagnetic torque in the above specified steady-state conditions for both the controllers is tabulated in Table IV.

The proposed ANFIS-HCC controller has lower torque ripples than PI-HCC at low load and low rpm, but it is comparable to the conventional PI-HCC-based FOC at higher rpm and higher load. Thus, experimentally validating the better performance of the proposed ANFIS-HCC controller in comparison with the conventional PI-HCC controller with superior torque characteristics with fewer ripples under steady-state conditions for various operating conditions.

Table IV. Torque Ripple under Steady State Operation.

Control Method	$750rpm - 3Nm$	$1500rpm - 6Nm$	$2250rpm - 9Nm$
PI-HCC	2. 05	0.53	0.31
ANFIS-HCC	0.80	0.40	0.22

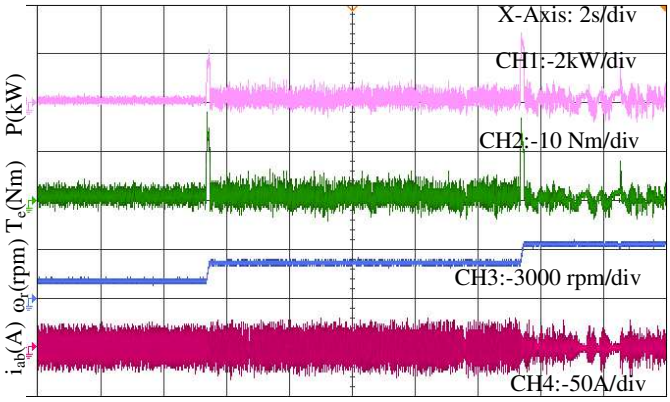


Fig. 14. Dynamic response of PMSM drive with PI-HCC based FOC for step-change in speed from 750 to 1500 to 2250 rpm under no load condition.

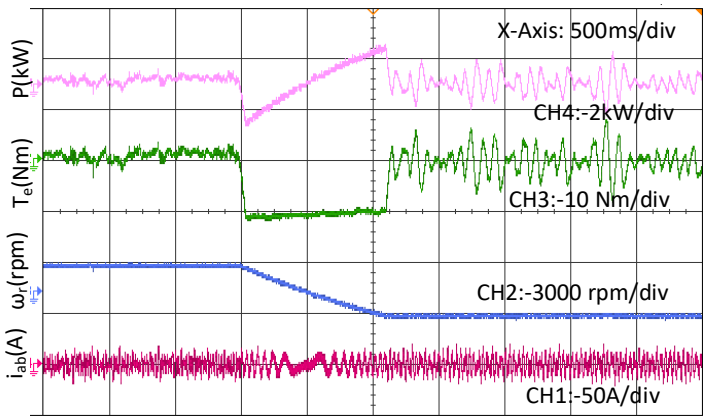


Fig. 15. Dynamic response of PMSM drive with PI-HCC based FOC for speed reversal from $+1500$ to -1500 rpm under no load condition.

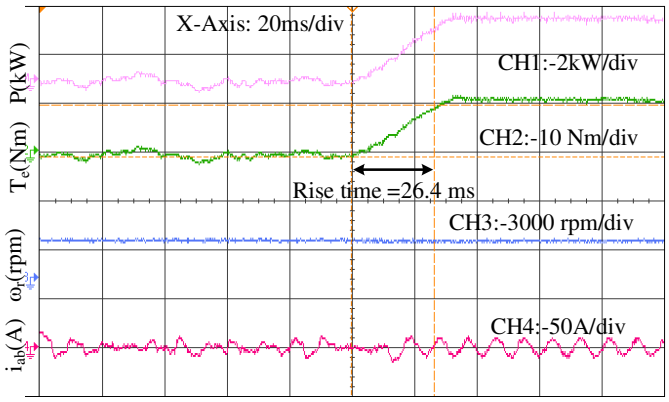


Fig. 16. Dynamic response of PMSM drive with PI-HCC based FOC for a step change in load from 0 to 9Nm at constant speed (2250rpm) operation.

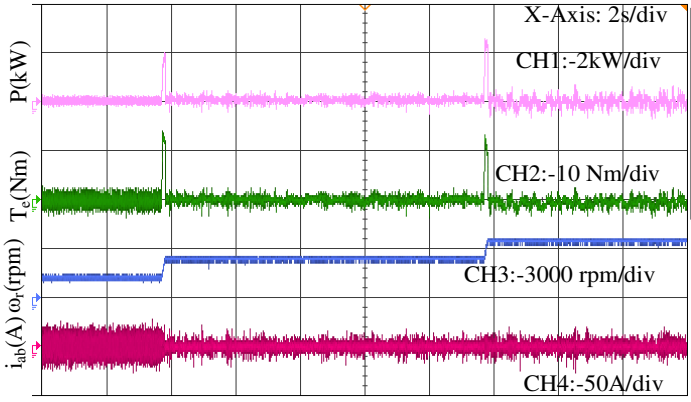


Fig. 17. Dynamic response of PMSM drive with ANFIS-HCC based FOC for step-change in speed from 750 to 1500 to 2250 rpm under no load condition.

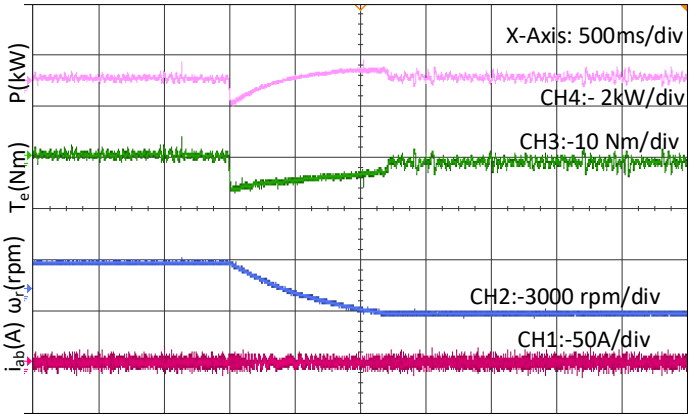


Fig. 18. Dynamic response of PMSM drive with ANFIS-HCC based FOC for speed reversal from $+1500$ to -1500rpm under no load condition.

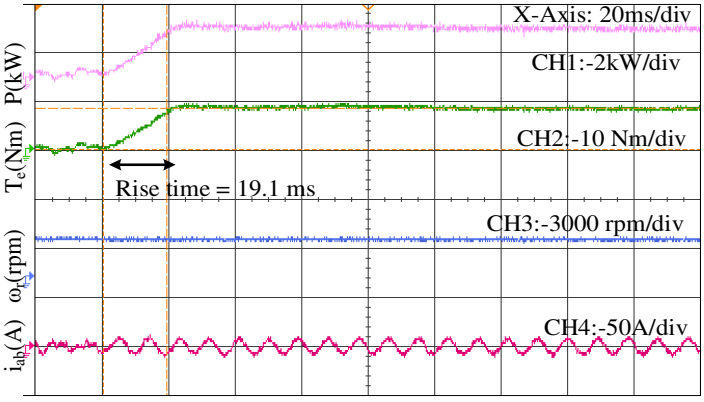


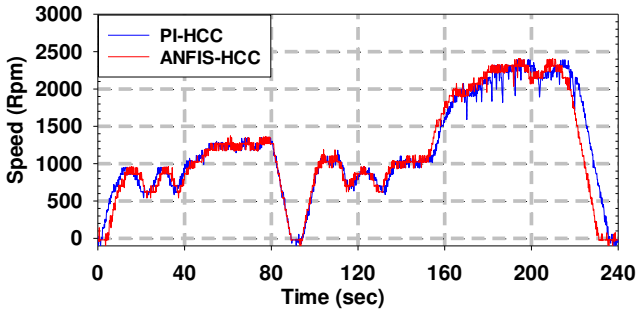
Fig. 19. Dynamic response of PMSM drive with ANFIS-HCC based FOC for a step change in load from 0 to 9Nm at constant speed (2250rpm) operation.

D. Dynamic Response of PMSM drive under variable speed/load torque conditions

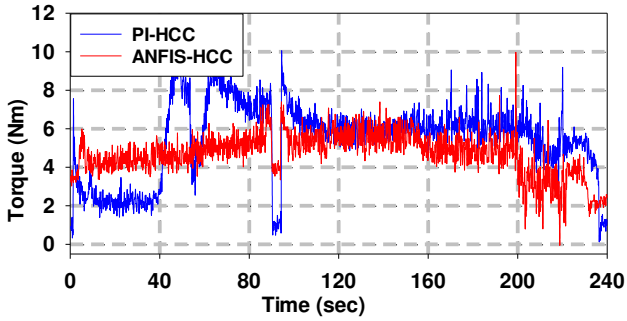
The dynamic response of the PMSM drives with conventional PI-HCC based FOC for step change in speed from 750 to 1500 to 2250rpm during forward motoring mode under no-load condition, for speed reversal from $+1500$ to $-1500\,pm$ under no load condition and step change in load from 0 to 9Nm at constant speed (2250 rpm) operation are shown in Figs 14 to 16 respectively. Figs.17 to 19 display the respective corresponding dynamic responses of the drive using the developed ANFIS- HCC for FOC of PMSM drive. It is observed that the proposed ANFIS-HCC-based controller reduces the torque and current ripples and provides smooth transition under step change in speed, step change in load, and change in motoring mode. The comparative analysis of the recovery time for the two controllers for a step change in load is given in Table V.

Table V. Comparative Analysis for step load change in load from 0 to 9Nm at 2250 rpm.

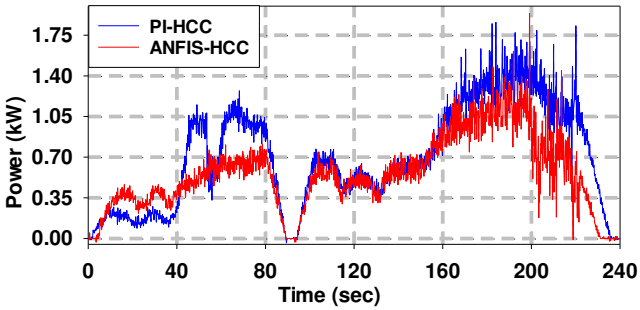
Parameter	PI-HCC	ANFIS-HCC
Rise-Time (ms)	26.4	19.1



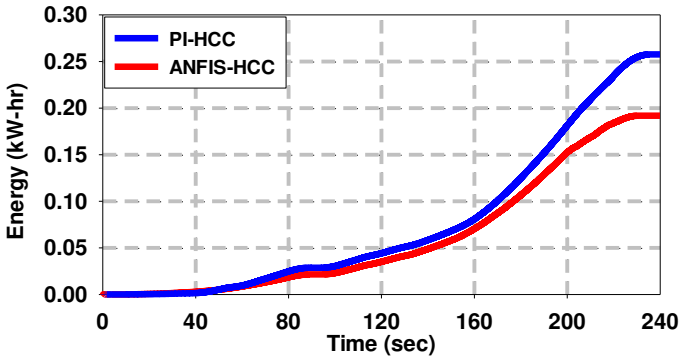
(a)



(b)



(c)



(d)

Fig. 20. Dynamic Response of the Laboratory Prototype for IM240 Drive Cycle. (a) Speed. (b) Torque. (c) Power. (d) Energy consumption.

E. Performance Validation using IM240 Drive Cycle.

Electrical motor drives used for EV applications have to withstand frequent changes in load torque and speed due to diverse traffic conditions, type of road surface, the gradient of the road, etc. It thus becomes imperative to test electric motor drives for EV applications under such varied operating conditions. The Inspection and Maintenance Driving Cycle (IM240), a $240sec$ test which represents a $3.1\,km$ driving track with a maximum speed of $91.2\ km/h$ and an average speed of $47.3\ km/h$ [26] is used to test the effectiveness and energy efficiency of the developed ANFIS-HCC based laboratory prototype.

Figs. 20(a) to 20(d) show the dynamic response of the ANFIS-HCC and PI-HCC-based FOC of the laboratory prototype when put through the IM240 Driving Cycle. It is observed that both the FOC models provide flawless speed tracking as shown in Fig. 20(a). However, during transients, there is a noticeable improvement in the generation of torque and a distinct lesser usage of power for developed ANFIS-HCC based FOC of PMSM Drive, as shown in Figs. 20(b) and 20(c) respectively. In addition, the proposed novel ANFIS-HCC-based drive utilizes less energy, with an energy consumption of 0.19kW-hr units compared to 0.26 kW-hr units for the PI-HCC-based drive, as shown in Fig. 20(d).

6. CONCLUSION

A novel ANFIS-HCC for FOC of PMSM Drive for EV application is designed and developed through simulation studies using MATLAB/Simulink. The proposed control scheme is implemented in a laboratory prototype using a DSP-based dSPACE1104 controller for experimental validation. The performance of the developed novel ANFIS-HCC-based FOC is compared with the conventional PI-HCC-based FOC for PMSM drive during starting, under steady state operation, during step change in load, step change in torque, and speed reversal. In addition, the suitability of the developed novel ANFIS-HCC-based FOC for PMSMS drive in EV application is validated by testing it using IM240 Drive Cycle, which depicts a typical traffic pattern. The modelling and experimental findings demonstrate the higher performance of the designed ANFIS-HCC-based controller under diverse operating situations.

As compared to the PI-HCC controller, the proposed ANFIS-HCC controller offers several advantages as specified below:

A) Under steady-state conditions, experimental findings show that the proposed ANFIS-HCC-based controller can reduce motor torque ripples by 60-65% at lower speeds and 10-20% at higher speeds.

B) A smoother dynamic response to a change in speed is observed for the proposed ANFIS-HCC controller. It exhibits up to 28% quicker torque recovery during step changes in load conditions.

C) The energy consumption for the IM240 drive cycle demonstrates that 27% less energy is required while using an ANFIS-HCC-based PMSM drive for EV applications.

It is thus established that the use of the proposed novel ANFIS-HCC-based FOC of PMSM Drive would result in a more efficient EV. The suggested drive can also be used for industrial applications like spindle drives [27] and precise positioning systems that require negligible ripples.

DECLARATION

We declare that the manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We know no conflicts of interest associated with this publication. I confirm that the manuscript has been read and approved for submission by all the named authors.

ACKNOWLEDGEMENT

This work is supported by Centre of Excellence, Electric Vehicle and Related Technologies, Delhi Technological University, Delhi

REFERENCES

[1] Z. Wang, T. W. Ching, S. Huang, H. Wang, and T. Xu, "Challenges Faced by Electric Vehicle Motors and Their Solutions," *IEEE Access*, vol. 9, pp. 5228–5249, 2021, doi: 10.1109/ACCESS.2020.3045716.

[2] C. T. Krasopoulos, M. E. Beniakar, and A. G. Kladas, "Velocity and Torque Limit Profile Optimization of Electric Vehicle Including Limited Overload," *IEEE Trans. Ind. Appl.*, vol. 53, no. 4, pp. 3907–3916, 2017, doi: 10.1109/TIA.2017.2680405.

[3] J. Yang and G. Zhu, "Stochastic Predictive Boundary Management for a Hybrid Powertrain," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4700–4713, 2016, doi: 10.1109/TVT.2015.2504325.

[4] Y. Zhang, W. Cao, S. McLoone, and J. Morrow, "Design and Flux-Weakening Control of an Interior Permanent Magnet Synchronous Motor for Electric Vehicles," *IEEE Trans. Appl. Supercond.*, vol. 26, no. 7, pp. 1–6, 2016, doi: 10.1109/TASC.2016.2594863.

[5] X. Sun, Z. Shi, G. Lei, Y. Guo, and J. Zhu, "Analysis and Design Optimization of a Permanent Magnet Synchronous Motor for a Campus Patrol Electric Vehicle," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10535–10544, 2019, doi: 10.1109/TVT.2019.2939794.

[6] Suryakant, M. Sreejeth, M. Singh, and A. K. Seth, "Minimization of torque ripples in PMSM drive using PI- resonant controller-based model predictive control," *Electr. Eng.*, vol. 105, no. 1, pp. 207–219, 2023, doi: 10.1007/s00202-022-01660-y.

[7] Z. Zhou, C. Xia, Y. Yan, Z. Wang, and T. Shi, "Torque Ripple Minimization of Predictive Torque Control for PMSM with Extended Control Set," *IEEE Trans. Ind. Electron.*, vol. 64, no. 9, pp. 6930–6939, 2017, doi: 10.1109/TIE.2017.2686320.

[8] H. Zhu, X. Xiao, and Y. Li, "Torque ripple reduction of the torque predictive control scheme for permanent-magnet synchronous motors," *IEEE Trans. Ind. Electron.*, vol. 59, no. 2, pp. 871–877, 2012, doi: 10.1109/TIE.2011.2157278.

[9] J. Lara, J. Xu, and A. Chandra, "Effects of Rotor Position Error in the Performance of Field-Oriented-Controlled PMSM Drives for Electric Vehicle Traction Applications," *IEEE Trans. Ind. Electron.*, vol. 63, no. 8, pp. 4738–4751, 2016, doi: 10.1109/TIE.2016.2549983.

[10] Q. Teng, "Nonlinear Function Integral Sliding Mode-Based Model Predictive Current Control for PMSM Drives With DC-Bus Voltage Observer," *CPSS Trans. Power Electron. Appl.*, vol. 7, no. 4, pp. 399–408, 2022, doi: 10.24295/cpsstpea.2022.00036.

[11] L. Yuan, Y. Jiang, L. Xiong, and P. Wang, "Sliding Mode Control Approach with Integrated Disturbance Observer for PMSM Speed System," *CES Trans. Electr. Mach. Syst.*, vol. 7, no. 1, pp. 118–127, 2023, doi: 10.30941/CESTEMS.2023.00009.

[12] C. Xia, C. Guo, and T. Shi, "A neural-network-identifier and fuzzy-controller-based algorithm for dynamic decoupling control of permanent-magnet spherical motor," *IEEE Trans. Ind. Electron.*, vol. 57, no. 8, pp. 2868–2878, 2010, doi: 10.1109/TIE.2009.2036030.

[13] H. A. G. Al-Kaf and K. B. Lee, "Low Complexity MPC-DSVPWM for Current Control of PMSM Using Neural Network Approach," *IEEE Access*, vol. 10, no. October, pp. 132596–132607, 2022, doi: 10.1109/ACCESS.2022.3230356.

[14] V. K. Arun Shankar, S. Umashankar, P. Sanjeevikumar, and S. Paramasivam, "Adaptive Neuro-Fuzzy Inference System (ANFIS) based Direct Torque Control of PMSM driven centrifugal pump," *Int. J. Renew. Energy Res.*, vol. 7, no. 3, pp. 1436–1447, 2017, doi: 10.20508/ijrer.v7i3.5885.g7177.

[15] H. N. Tran and J. W. Jeon, "Robust Speed Controller Using Dual Adaptive Sliding Mode Control (DA-SMC) Method for PMSM Drives," *IEEE Access*, vol. 11, no. June, pp. 63261–63270, 2023, doi: 10.1109/ACCESS.2023.3288124.

[16] Y. Feng, H. Zhao, M. Zhao, and H. Chen, "A Feedback Linearization Control Scheme Based on Direct Torque Control for Permanent Magnet Synchronous Motor," *Chinese Control Conf. CCC*, vol. 2018-July, pp. 87–92, 2018, doi: 10.23919/ChiCC.2018.8483803.

[17] Z. H. Liu, J. Nie, H. L. Wei, L. Chen, X. H. Li, and M. Y. Lv, "Switched PI Control Based MRAS for Sensorless Control of PMSM Drives Using Fuzzy-Logic-Controller," *IEEE Open J. Power Electron.*, vol. 3, no. May, pp. 368–381, 2022, doi: 10.1109/OJPEL.2022.3182053.

[18] A. V. Sant, "PM synchronous motor speed control using hybrid fuzzy-PI with novel switching functions," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 4672–4675, 2009, doi: 10.1109/TMAG.2009.2022191.

[19] M. A. Islam *et al.*, "Modeling and Performance Evaluation of ANFIS Controller-Based Bidirectional Power Management Scheme in Plug-In Electric Vehicles Integrated with Electric Grid," *IEEE Access*, vol. 9, pp. 166762–166780, 2021, doi: 10.1109/ACCESS.2021.3135190.

[20] K. N. Sujatha and K. Vaisakh, "Implementation of Adaptive Neuro Fuzzy Inference System in Speed Control of Induction Motor Drives," *J. Intell. Learn. Syst. Appl.*, vol. 02, no. 02, pp. 110–118, 2010, doi: 10.4236/jilsa.2010.22014.

[21] J. Li, W. Song, H. Yue, N. Sun, C. Ma, and R. Feng, "An Improved MPC With Reduced CMV and Current Distortion for PMSM Drives Under Variable DC-Bus Voltage Condition in Electric Vehicles," *IEEE Trans. Power Electron.*, vol. 38, no. 4, pp. 5167–5177, 2023, doi: 10.1109/TPEL.2022.3227436.

[22] T. Lee and K. Nam, "Torque Control Based Speed Synchronization for Two-Speed Gear Electric Vehicle," *IEEE Access*, vol. 9, pp. 153518–153527, 2021, doi: 10.1109/ACCESS.2021.3127368.

[23] C. Lu, S. Ferrari, and G. Pellegrino, "Two Design Procedures for PM Synchronous Machines for Electric Powertrains," *IEEE Trans. Transp. Electrif.*, vol. 3, no. 1, pp. 98–107, 2017, doi: 10.1109/TTE.2016.2646738.

[24] J. R. Jang, "Neuro-Fuzzy Modeling," vol. 83, no. 3, 1995.

[25] J. R. Jang, "ANFIS : Adap tive-Ne twork-Based Fuzzy Inference System," vol. 23, no. 3, 1993.

[26] "Dynamometer drive schedules", [online] Available: https://www.epa.gov/vehicle-and-fuel-emissions-testing.

[27] P H. Zhao, S. Yu and F. Sun, "Harmonic Suppression and Torque Ripple Reduction of a High-Speed Permanent Magnet Spindle Motor," in IEEE Access, vol. 9, pp. 51695-51702, 2021, doi: 10.1109/ACCESS.2021.3070010..

Physica Scripta

ACCEPTED MANUSCRIPT

Physics Based Numerical Model of a Nanoscale Dielectric Modulated Step Graded Germanium Source Biotube FET Sensor: Modelling and Simulation

To cite this article before publication: Amit Das *et al* 2023 *Phys. Scr.* in press https://doi.org/10.1088/1402-4896/acf4c9

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is © **2023 IOP Publishing Ltd**.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript will be available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by-nc-nd/3.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 14.139.251.98 on 04/09/2023 at 07:42

Physics Based Numerical Model of a Nanoscale Dielectric Modulated Step Graded Germanium Source Biotube FET Sensor: Modelling and Simulation

Amit Das[1,*], Sonam Rewari[2,*], Binod Kumar Kanaujia[1], S.S. Deswal[3] and R.S. Gupta[4]

[1] School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi-110067, India
[2] Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi-110042, India
[3] Department of Electrical and Electronics Engineering, Maharaja Agrasen Institute of Technology, New Delhi-110086, India
[4] Department of Electronics and Communication Engineering, Maharaja Agrasen Institute of Technology, New Delhi-110086, India
[*] Authors to whom any correspondence should be addressed.

E-mail: amitofficial7492@gmail.com and rewarisonam@gmail.com

Keywords: biotube, gate all around, MOSFET, MOSFET biosensors, SILVACO TCAD, step-graded doping, surrounding gate MOSFET

Abstract

This paper proposes a novel dielectric modulated step-graded germanium source biotube FET for label-free biosensing applications. Its integrated structure and unique design combine the benefits of the gate stack, germanium source, triple-gate architecture, and a step-graded biotube channel, resulting in superior performance over existing biosensors. A compact two-dimensional analytical model for channel potential, drain current, threshold voltage, and subthreshold swing has been formulated and agrees well with the simulated results. The comprehensive investigation of different device parameters, including doping and bias, offers valuable insights into optimizing the biosensor's performance. The proposed biosensor exhibits remarkable sensitivity, achieving up to 263 mV and 1495.52 nA for certain biomolecules, which has been validated by a compact analytical model and simulations performed on the SILVACO TCAD simulator. Several parameters are employed to assess the biosensor's effectiveness: threshold voltage, I_{ON}/I_{OFF} ratio, subthreshold swing, off-current, peak trans-conductance, and on-current. Furthermore, the biotube channel design enables lightweight and cost-efficient biosensors, enhancing the biosensor's practicality. This work also includes an analysis of the effect of temperature on the biosensor's performance and characteristics, providing insights into practical applications. High sensitivity of the biosensor signifies a significant advancement in biosensing technology, suggesting a wide range of potential applications in biomedical field.

List of abbreviations and symbols

Abbreviation	Description
CMOS	Complementary Metal Oxide Semiconductor
DM	Dielectric Modulated
DM-FET	Dielectric Modulated-Field Effect Transistor
DM-SGGS-BTFS	DM-Surrounding Gate Germanium Source-Biotube FET Sensor
FET	Field Effect Transistor
FTSGP	Forward Triple Step Graded Profile
IIFET	Impact Ionization Field Effect Transistor
g_d	Output Conductance
g_m	Transconductance
g_{mp}	Peak Transconductance
I_{ON}	On-Current
I_{OFF}	Off-Current
I_{ON}/I_{OFF}	Current Ratio
K_{bio}	Dielectric Constant of Biomolecules
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
N_f	Charge Density of Biomolecules
RTSGP	Reverse Triple Step Graded Profile
SCE	Short Channel Effects
SG-MOSFET	Surrounding Gate-MOSFET
S_{FCR}	Fractional Sensitivity in I_{ON}/I_{OFF} Ratio
S_{gm}	Transconductance Sensitivity

Accepted Manuscript

S_{gd}	Output Conductance Sensitivity
S_{gmp}	Peak Transconductance Sensitivity
S_{IDS}	Drain Current Sensitivity
S_M	Sensitivity with respect to sensing metric 'M'
S_{SS}	Subthreshold Swing Sensitivity
S_{Vt}	Threshold Voltage Sensitivity
TCAD	Technology Computer Aided Design
TFET	Tunnel Field Effect Transistor
V_{GS}	Gate Voltage
V_{DS}	Drain Voltage

1. Introduction

The exponential and rapid development of technology has led to the progressive evolution of Field Effect Transistor (FET) devices [1–4]. Field effect transistors are effectively used in sensing applications due to scalability and high sensitivity. Sensing technologies have evolved over the years, leveraging advancements in electronics and miniaturization to enhance their capabilities. From simple temperature and pressure sensors to sophisticated devices like pH/biomolecules/gas sensors, the range of sensing applications has expanded rapidly. Most of the FET sensors utilize the basic principles of dielectric modulation to detect and measure various physical parameters. To enhance the selectivity of these dielectric modulated FET based biosensors, the FET device is integrated with a sensing element, which could be a specific material, a biomolecule, or an enzyme. The development of FET sensing and biosensors has opened up new possibilities for real-time, label-free, and portable sensing devices, with the potential to transform healthcare, agriculture, and environmental monitoring.

FET based biosensors generally utilize the principle of dielectric modulation for label-free detection of biochemical species. Dielectric modulated FET (DM-FET) biosensors have gained attention in recent years for their potential applications in biomedical field [5,6]. The main advantage of a DM-FET based biosensor is its ability to detect biomolecules without the need for labeling (label-free detection). Generally, when biomolecules are trapped within the embedded cavity (without any bioreceptor layer), the characteristics of FET undergo changes that enable the detection of biomolecules by analyzing the relative change in the sensing metric (threshold voltage/drain current/current ratio/subthreshold swing). However, a significant drawback of this type of detection is its poor selectivity [6]. To address this limitation, the selectivity of the DM-FET biosensor can be greatly enhanced by incorporating an additional bioreceptor layer, which is functionalized on the gate oxide layer and specifically binds to the target biomolecule (biotarget). The incorporation of an additional bioreceptor layer significantly improves the selectivity but could increase the complexity of the fabrication process in the nanoscale regime and may lead to a decrease in sensitivity (considering the relative change in the value of the sensing metric). Various biotarget - bioreceptor systems have been employed in FET biosensors, as reported in the existing literature [4,7–10]. Examples include the interaction between Streptavidin and Biotin, the utilization of Iris antibody (anti-Iris) for binding to the Iris antigen, and the Avian influenza antibody (anti-AI) targeting the Avian influenza surface antigen (AIa).

In real scenario, DM-FET biosensor utilizes the concept of dielectric modulation to detect the presence or absence of biomolecules. The gate oxide layer is functionalized with a bioreceptor capable of specifically binding the desired biomolecule, known as the biotarget. The interaction between the biotarget and bioreceptor leads to alterations in the electrical characteristics of the device (FET), including electric potential, conduction current, and electron concentration in the channel. The observed changes in electrical properties (characteristics) are similar to the changes induced by the application of an external gate voltage (gate/drain voltage). The degree and extent of changes in the sensing metrics following the binding of biomolecules to the bioreceptor can be utilized for biomolecule detection. Therefore the sensitivity of a DM-FET biosensor can be defined as the change in the biosensor's output, which corresponds to variations in the concentration/charge density of analyte, serves as an indicator for detecting the presence or absence of biomolecules. This detection relies on properties such as the dielectric constant (or permittivity) as well as, charge density of the biomolecules [5–7]. However, DM-FET biosensors also have some disadvantages such as sensitivity to operating conditions and environmental conditions (temperature, humidity, high/low frequency noise etc); manufacturing complexity achieving a high level of accuracy and precision in creating nanometer-scale cavities with desired shape and perfection); selectivity issues (limitation in multi-analyte detection: simultaneous detection of multiple analytes in a single complex sample is challenging and non-specific binding: leads to false-positive output); and stability issues along with the reliability issues affecting the long-term performance (various environmental factors and aging effects can lead to changes in the properties of the dielectric layer, which can impact the biosensor's electrical response; and furthermore, the presence of different biomolecules trapped within the cavity can gradually cause tear over extended periods).

FET has specifically three primary variants which are utilized in biosensing applications: Metal Oxide Semiconductor Field Effect Transistor (MOSFET), Tunnel Field Effect Transistor (TFET), and Impact Ionization Field Effect Transistor (IIFET). TFET and IIFET are the newer variants, whereas MOSFET is one of the oldest and utilitarian variants in the FET family. The rapid progress in nano-bioelectronics has opened up potential applications for MOSFET

Accepted Manuscript

in label-free electrical detection of biomolecules. MOSFET has gained attention due to their high sensitivity, compatibility with CMOS technology, and the ease of fabricating biosensing systems onto integrated circuits. The development of TFET based biosensors has emerged as a promising avenue for sensitive and low-power biosensing applications [11–13]. TFET, which operate on the principle of quantum tunneling, offer distinct advantages over traditional MOSFET, such as lower power consumption and reduced leakage currents. These unique characteristics make TFET well-suited for biosensing, where high current sensitivity (subthreshold region) and minimal energy consumption are critical. Dwivedi et al. [14] has addressed the location-dependent sensitivity degradation in conventional TFET by introducing a pocket TFET biosensor, which combines lateral and vertical tunneling in a single architecture. Dwivedi et al. [7] has compared tunneling and accumulation mode p-type transistor architectures as dielectric modulated biosensors, showing that while TFET achieve higher sensitivity near the source-channel junction, accumulation mode transistors exhibit higher sensitivity for biomolecules located away from the junction. While TFET and IIFET offers unique advantages such as lower power consumption or impact ionization-induced amplification, MOSFET currently provide better overall performance, reliability, and compatibility with existing technology. Conventionally, MOSFET is often considered superior to TFET and IIFET for several reasons:-

i. **Compatibility with existing and latest technology:** MOSFET has been widely used in the semiconductor industry for many years and are compatible with standard CMOS fabrication processes. The manufacturing process for MOSFET is well-established/known, resulting in high yield and reliable production. In contrast, TFET and IIFET are relatively new variants (technologies) and their fabrication processes are still being developed, making them potentially less accessible and reliable (industrial concern/drawback). Furthermore, the issue of random dopant fluctuations is more prevalent in TFET compared to MOSFET. This can be attributed to the conventional pin-structure of TFET, which necessitates high abrupt doping and the use of different materials in the source and drain regions (increasing fabrication complexity) [15,16].

ii. **No ambipolar conduction:** The issue of ambipolarity in TFET makes them more vulnerable to temperature fluctuations, resulting in a degradation of their noise performance. Additionally, in the context of sensing applications, a noteworthy drawback emerges as various sensing metrics become highly temperature dependent. This means that even a slight change in temperature can cause a significant shift in the value of the sensing metric, ultimately leading to undesirable temperature-dependent sensitivity [15].

iii. **Higher on-state current:** MOSFET generally exhibit higher on-current than TFET resulting in a slightly better transconductance and current ratio (comparatively, more suitable in amplifiers, and digital applications). Additionally, transconductance and current ratio can be valuable sensing metrics when utilizing MOSFET for different sensing applications.

iv. **Performance and speed:** MOSFET typically offer higher performance and comparable speed compared to TFET and IIFET. MOSFET has been extensively optimized over the years, allowing for high-speed operation and efficient switching applications. TFET and IIFET, while offering potential advantages such as lower power consumption and lower subthreshold swing, often suffer from reduced performance [15,17–20] in terms of speed (tunneling phenomenon causing delay intrinsic delay in charge carrier transport), noise performance (ambipolar behavior) and lower on-state current driving capability.

v. **Scalability and integration:** MOSFET has a well-established scaling roadmap, enabling the fabrication of smaller and more integrated devices. They can be readily integrated with other electronic components, allowing for the development of complex systems and sensor arrays. TFET and IIFET (being new emerging devices) may face challenges in terms of scalability and integration due to their unique device structures and associated fabrication complexities [5,21].

vi. **Wide range of applications:** MOSFET has been extensively studied and applied in various fields, including electronics, telecommunications, and biosensing. The versatility, scalability and compatibility with existing technology make them suitable for a wide range of applications. TFET and IIFET, being relatively new technologies (in terms of usability), are still exploring their potential applications (sensing and biomedical field) and may have limited practical implementations.

The exposition of FET based biosensors for sensing nucleic acids [22], pH levels [23,24], proteins [25], and DNA [26] has been evaluated and shown in the past. Buitrago et al. [27] has experimentally demonstrated a silicon nanowire-based FET biosensor. Many non-planar variants of surrounding gate MOSFET biosensors have been reported in the recent decade. Pratap et al. [28] has proposed a novel MOSFET-based biosensor that utilizes the concept of a junctionless transistor [29]. Chakraborty et al. [30] reported a variant of surrounding gate MOSFET-based biosensor and has shown the effect of gate oxide stacking on the sensitivity. High-k dielectric has numerous advantages [31–33], but direct deposition of HfO_2 over a silicon substrate is complex due to stability issues and can also degrade the performance of the biosensor. Therefore, HfO_2 is deposited over a thin layer of SiO_2 [34,35]. The literature survey reveals that channel doping [36,37] and source doping [38,39] must be kept low and high, respectively, to obtain superior characteristics. Therefore, a detailed study of doping-dependent sensitivity analysis becomes inevitable when investigating the performance of any FET-based biosensor.

Accepted Manuscript

MOSFET based biosensors [6] have gained attention due to their scalability and high sensitivity [30]. The key principle behind the biosensing ability of MOSFET-based biosensors is the relative change in the value of device metrics in the presence of biomolecules. Dielectric modulation enables the gate to control the flow of charge carriers. When biomolecules are immobilized inside the cavity, they change the net gate oxide capacitance, which in turn alters the potential and field distribution. This modulation affects device characteristics, such as subthreshold slope, off-current, and threshold voltage. The dielectric modulation of these parameters is the primary reason behind the biosensing capability of MOSFET. MOSFET-based biosensors are capable of label-free detection, which utilizes different molecular properties to sense biomolecules [40,41].

Optimization of device parameters can result in a highly sensitive biosensor. Goel et al. [42] revealed that the sensitivity of biosensors increases with the increase in the number of gates used. The proper choice of source material can substantially improve the sensitivity of the biosensor. Wu et al. [43] and Brunco et al. [44] have discussed the concept of germanium MOSFET in the past decade. Saha et al. [45] reported a germanium source-based TFET for label-free detection of biomolecules. Germanium offers more intrinsic charge carriers than silicon, which participates in the conduction process [46] and are involved in the biosensing action. The relative change in the various sensing metrics will be higher when the numbers of charge carriers are high. Hence, the sensitivity of the biosensor is higher in germanium-based biosensors. Using a graded channel instead of a non-graded channel improves the device's performance [47,48]. Recently, Singh et al. [49] reported a label-free FET biosensor that uses a graded channel. Although germanium-based biosensors are highly sensitive, they have some flaws and drawbacks. Germanium source-based biosensors are costlier than their silicon-based counterparts from an economic perspective, and their use is limited due to the limited availability of fabrication technology. Additionally, germanium has a lower operating temperature range than silicon, which limits the range of temperatures at which it can effectively operate.

This paper comprehensively investigates a dielectric-modulated step-graded germanium-source biotube FET sensor (DM-SGGS-BTFS). It shows enhanced sensitivity due to source and channel engineering, which involves using a germanium source and a graded-doped biotube filled with air [50]. Both physics-based modeling and simulation-based investigation are crucial to optimize the different parameters for designing an ultra-sensitive biosensor. The key novelty of this research work lies in two aspects: the structural innovation of the biotube sensor and the comprehensive investigation through analytical modeling, considering different sensing metrics and the impact of temperature. The structural novelty involves an integrated design of the surrounding gate MOSFET, which incorporates gate engineering (triple gate architecture), source engineering (germanium source), channel engineering (tubular channel with step-graded doping profile), and gate oxide engineering (oxide stacking: HfO_2+SiO_2) techniques. The biosensing performance of DM-SGGS-BTFS has been effectively analyzed using multiple key sensing metrics such as threshold voltage, subthreshold swing, I_{ON}/I_{OFF} ratio, drain current, peak transconductance, and peak output conductance. Additionally, the developed analytical model takes into account the influence of temperature on various analog performance parameters and sensing parameters of the proposed biosensor. One of the primary objectives of this paper is to present an analytical model of DM-SGGS-BTFS to investigate its biosensing performance and study the critical insights into the sensitivity dependency on different design parameters. To make this analysis more practical and realistic, the effect of partially filled cavities on sensitivity has also been considered. The analytical model of the biosensor is based on the non-linear parabolic solution of the 2D Poisson's equation [30]. The potential across the channel is expressed in terms of channel-surface potential [51,52]. The channel potential, threshold voltage, subthreshold swing, and drain current have been analytically derived, and the results show a remarkable agreement with the simulated results.

2. Device structure and software specifications

Figures 1(a) and 1(b) show the 3D cylindrical structure and a cross-sectional view, respectively, of the DM-SGGS-BTFS. A layer of HfO_2 is formed over the thin layer of SiO_2 to form the gate stack. The biotube has an inner radius (a) of 2 nm and an outer radius (b) of 10 nm. The biosensor considers a dual-sided cavity, which has the advantage of better fill-in factor probability than the single-sided cavity. Additionally, the asymmetric cavity has been preferred since the source-end open cavity is comparatively more sensitive to biomolecules than the drain-end open cavity [53]. The work function of the gate in the multi-gated structures can also affect the sensitivity of the biosensor [54]. Kumari et al. [54] in her work mentioned that the sensitivity obtained is highest if the work functions of the gate near to drain is smaller than that of the source ($\phi_{M1}>\phi_{M3}$). Therefore, we have extended the same results for the proposed biosensor in our investigation. All the structural parameters of the biosensor are listed in table 1. Table 2 lists the lattice parameters of silicon and germanium in a tabular form. This paper primarily focuses on the biosensing performance of DM-SGGS-BTFS, and biosensing investigation requires changes in the value of sensing metrics. The net electric field across the channel can be resolved into two components: the lateral electric field (across the radial direction) and the horizontal electric field (across the z-direction). The vertical electric field is the primary factor that controls the sensitivity and affects the biosensor's performance. The vertical electrical field is changing due to the immobilization of biomolecules inside the cavity which ultimately changes the gate oxide capacitance [3]. The sensitivity of the biosensor is calculated based on the relative change in the value of the sensing metric. The horizontal electric field (doping dependent) is almost constant since the

Accepted Manuscript

doping of the source, channel, and drain are constant (N_S=N_D=10^{19}/cm^3 and $N_{Cha,1}$=10^{10}/cm^3, $N_{Cha,2}$=10^{14}/cm^3, $N_{Cha,3}$=10^{18}/cm^3) [6,55], so it has a negligible effect on the sensitivity.

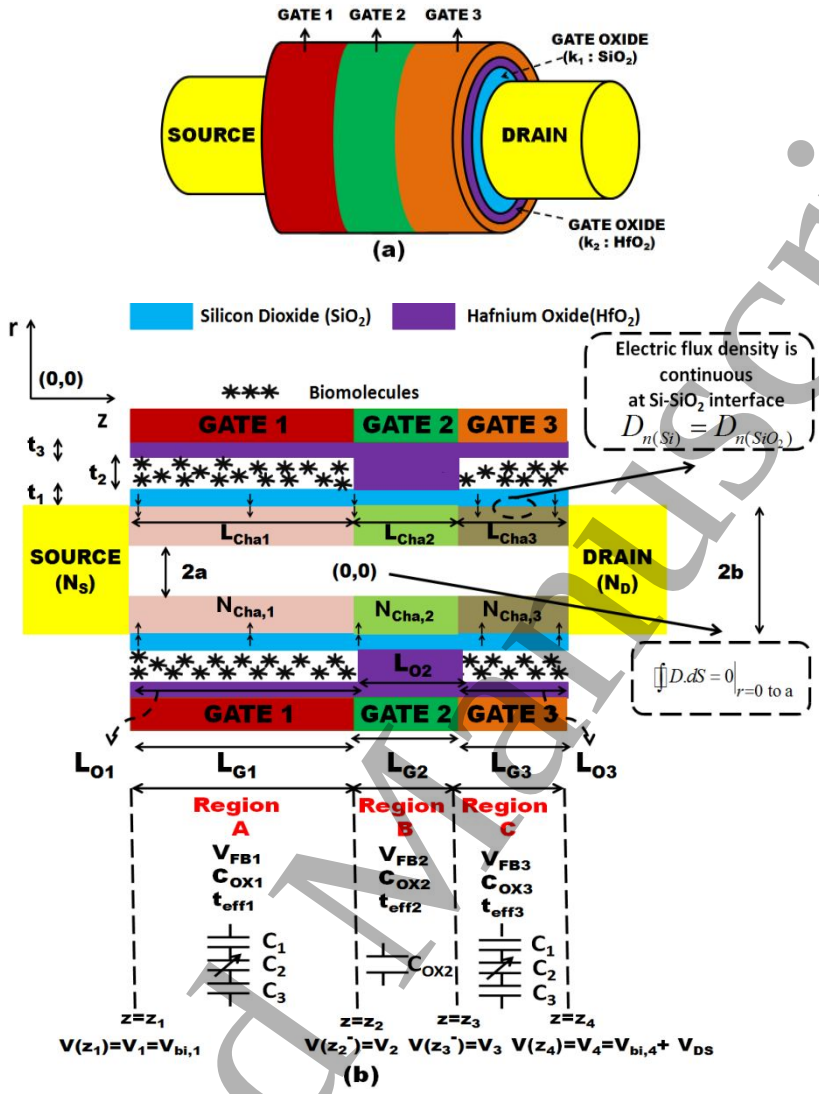


Figure 1. (a) Cylindrical (3D), and **(b)** cut-plane (2D) view of DM-SGGS-BTFS.

Table 1. Structural parameters used in the simulation of DM-SGGS-BTFS.

Parameter		Value		Parameter			Value	
Doping	Drain (N_D)	10^{19}/cm^3	Silicon	Gate length & Work function	L_{G1}	ϕ_{M1}	30 nm	5.10 eV [56]
	Channel (N_{Cha})	10^{10}-10^{18}/cm^3	Silicon		L_{G2}	ϕ_{M2}	15 nm	4.53 eV [56]
	Source (N_S)	10^{19}/cm^3	Germanium		L_{G3}	ϕ_{M3}	15 nm	4.10 eV [57]
Oxide thickness (t_{OX})		t_{O1}= t_{O3}=1nm	t_{O2}=6nm	Source/Drain length			20 nm	
Oxide length (L_{TOT})		L_{O1}=30nm	L_{O2}=L_{O3}=15nm	Channel radius (a & b)			a=2 nm	b=10nm
Cavity length (L_{CAV})		45 nm		Cavity thickness (t_{CAV})			6 nm	
n_i (T=300K)		Si: 1.5*10^{10}/cm^3	Ge: 2.5*10^{13}/cm^3	Boltzmann Constant (K_B)			1.38*10^{-23}J/K	
Biomolecules (K_{bio})		B_1: Streptavidin (K_{bio}=2.1) [42]		B_2 : Protein (K_{bio}=2.5) [28]			B_3 : Biotin (K_{bio}=2.63) [58]	
		B_4 : ChOx (K_{bio}=3.3) [28]		B_5 : APTES (K_{bio}=3.57) [30]			B_6 :Hydroprotein (K_{bio}=5) [59]	
		B_7 : Keratin (K_{bio}=8) [49]		B_8 : DNA (K_{bio}=5 & N_f=-1e10/cm^2 — -1e12/cm^2) [60,61]				
Physical models		BGN – Bandgap narrowing due to high doping						
		Boltzmann – Model carrier statistics		Lombardi CVT – Basic model for non planar MOSFET				
		Auger – To incorporate effects of recombination at high charge densities		FLDMOB – Model velocity saturation at high doping and temperature				
		CONMOB – Model concentration-dependent mobility of charge carriers		SRH – Recombination model used in surrounding gate MOSFET				

Accepted Manuscript

Table 2. Lattice parameters.

Lattice Parameters	Silicon	Germanium
Atomic number	14	32
Atomic weight	28.086 u	72.59 u
Density	2.33 g/cm^3	5.32 g/cm^3
Crystal structure	Diamond cubic (FCC)	Diamond cubic (FCC)
Atoms	5*10^{22} cm^{-3}	4.42*10^{22} cm^{-3}
Lattice constant	5.429 A^o	5.657 A^o
Nearest neighbour distance	0.235 nm	0.211 nm
Atomic radius	0.132 nm	0.137 nm
Nature of band gap	Indirect	Indirect
Dielectric constant	11.9	16
Energy gap	1.12 eV	0.67 eV
Electron affinity	4.05 eV	4 eV
Intrinsic carrier concentration	1.5*10^{10} cm^{-3}	2.5*10^{13} cm^{-3}
Intrinsic Debye length	24 μm	0.68 μm
Intrinsic resistivity	2.3*10^5 Ω-cm	45 Ω-cm
Effective DOS in CB (N_C)	2.8*10^{19} cm^{-3}	1.04*10^{19} cm^{-3}
Effective DOS in VB (N_V)	1.04*10^{19} cm^{-3}	6.0*10^{18} cm^{-3}
Melting point	1415^o C	936^o C
Electronic configuration	1s^2 2s^2 2p^6 3s^2 3p^2	1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^2

Figure 2(a) shows the overall fabrication flowchart [62–64] of DM-SGGS-BTFS [65], while figure 2(b) shows the comparison of its sensitivity with an existing FET-based biosensor [28]. The calibration of the proposed device with the experimental work reported by Choi et al. [66] is shown in figure 2(c). To better understand the improved biosensing action of the proposed biosensor, table 3 lists the structural parameters of a conventional surrounding gate MOSFET, and table 4 shows a sensitivity comparison of DM-SGGS-BTFS with a conventional surrounding gate MOSFET-based biosensor. The basic parameters and structure of conventional MOSFET and DM-SGGS-BTFS were kept the same to facilitate the comparison, and sensitivity was obtained by calculating the relative change (S_M) or the fractional change (S_{FM}) [6] where, M denotes the metric of the biosensor. Different sensing metrics have been considered in the analysis for a more realistic and reasonable comparison. Also, the fractional sensitivity improvement in the current ratio can be seen in the proposed structure, which can also be used as a potential sensing metric. It can be seen that the sensitivity improves significantly due to the structural engineering in DM-SGGS-BTFS when compared to a conventional MOSFET. Silicon and germanium have different mobility characteristics for electrons and holes, which is one of the primary reasons behind the different sensitivity characteristics. To be more precise, the mobility of the conducting electrons is approximately twice as high, while the mobility of holes is approximately four times higher in germanium compared to silicon [7]. The high mobility in germanium can be attributed to the lower effective mass of its electrons and holes. This mobility enhancement leads to an improvement in the subthreshold characteristics. Therefore, for the same range of gate or drain voltage, a biosensor utilizing a germanium source (Ge-source) will exhibit a comparatively greater variation (swing) in current compared to one utilizing a silicon source. This means DM-SGGS-BTFS utilizing a germanium source will exhibit a larger variation in sensing metrics like subthreshold swing and threshold voltage, increasing its sensitivity and making it more sensitive to biomolecules when compared to a biosensor utilizing a silicon source.

DM-SGGS-BTFS is a label-free biosensor that uses the molecular and physical properties of biomolecules to detect and sense them [30,67–70]. Kim et al. [71] has practically demonstrated the entrapment of streptavidin (size ~5 nm) inside a cavity. Different biomolecules are variable in size ranging from few micrometers to few attometers [28]. A dual sided asymmetric cavity has been considered which facilitates a high fill-in factor probability along with enhanced sensitivity [72]. This work considered the different biomolecules (K_{bio}=2-8 and size<6 nm) which can be easily localized and entrapped inside the cavity (size: 30*6 nm^2) of DM-SGGS-BTFS. The orientation of biomolecules also affects the localization, i.e. even few biomolecules whose size is greater than 6 nm (less than 30 nm) can be localized inside the cavity in horizontal orientation.

The numerical simulator used is SILVACO TCAD [73] which uses the NEWTON GUMMEL method to solve the complex coupled differential equations. Different models that have been used are mentioned in table 1. For the simulations performed on the TCAD simulator, the cavity is presumed to be filled with a material having some finite dielectric constant ($K_{bio}\neq1$) when the entire cavity is pervaded with different biomolecules. When the cavity is empty (no biomolecules are present), it is presumed to be filled with air (K_{air}=1). The asymmetric cavity provides comparatively more sensitivity when compared to symmetric cavity. Various biomolecules have been considered in this work, which solidifies the biosensing investigation of DM-SGGS-BTFS. All the test biomolecules are mentioned in table 1 and have been used to analyze the sensitivity of biosensors in the past [28,30,42].

Accepted Manuscript

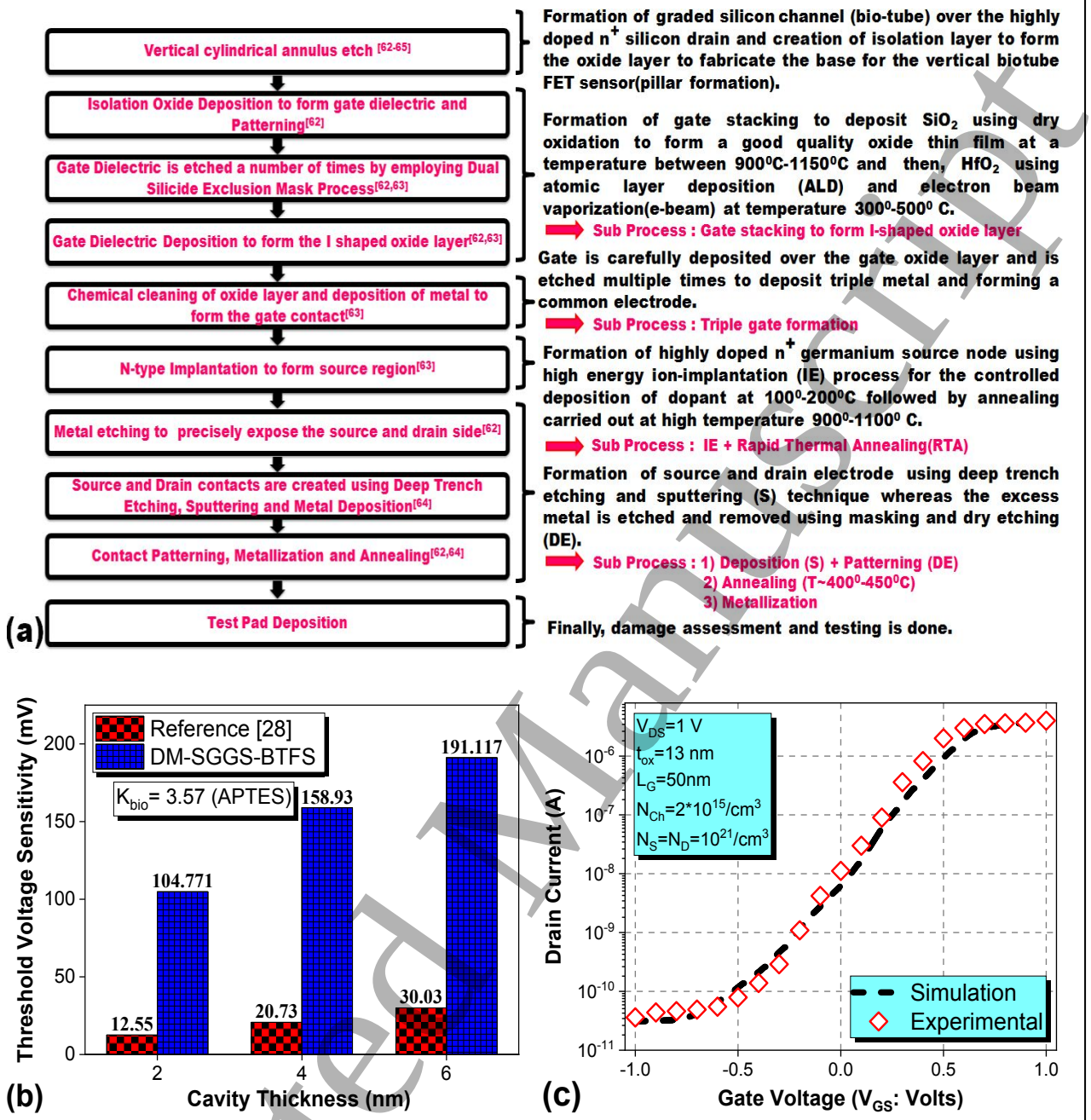


Figure 2. (a) Flow chart illustrating the basic fabrication steps [74], **(b)** sensitivity comparison and **(c)** calibration with experimental work [66] in DM-SGGS-BTFS.

Table 3. Parameters description of DM-SGGS-BTFS and conventional MOSFET.

Parameters	Conventional MOSFET	Our Work
Source/Drain doping	10^{19} cm^{-3}	10^{19} cm^{-3}
Channel doping	10^{10} cm^{-3}	$N_{Cha,1}=10^{10}$ cm^{-3} / $N_{Cha,2}=10^{14}$ cm^{-3} / $N_{Cha,3}=10^{18}cm^{-3}$
Work function	4.8 eV	ϕ_{M1}=5.1 eV / ϕ_{M2}= 4.53 eV / ϕ_{M3}=4.1 eV
Gate oxide	SiO_2 (8nm)	SiO_2 (1nm) + HfO_2 (7nm)
Channel length	60 nm	60 nm
Cavity length	30 nm + 15 nm	30 nm + 15 nm
Cavity thickness	6 nm	6 nm
Channel radius	10 nm	10 nm
Source material	Silicon	Germanium
L_{O1}=30nm , L_{O2}=15nm , L_{O3}=15nm		$L_{Cha,1}$=30nm , $L_{Cha,2}$=15nm , $L_{Cha,3}$=15nm
L_{G1}=30nm , L_{G2}=15nm , L_{G3}=15nm		

Accepted Manuscript

Table 4. Sensitivity comparison of DM-SGGS-BTFS with a conventional SG-MOSFET.

Biosensor	**Sensitivity**	$S_M = \lvert M(without_biomolecules) - M(with_biomolecules) \rvert$ [28]						
		M : Threshold Voltage (S_{Vt} : mV)		M : Subthreshold Swing (S_{SS} : mV/decade)			M : Off Current (S_{IOFF} : fA)	
		$S_{FCR} = \left\lvert \frac{CR(with_biomolecules)}{CR(without_biomolecules)} \right\rvert$		$CR = \frac{I_{ON}}{I_{OFF}}$ [70]			I_{ON} : On Current I_{OFF} : Off Current CR : Current Ratio (10^7)	
Neutral biomolecules		**B_1**	**B_2**	**B_3**	**B_4**	**B_5**	**B_6**	**B_7**
Conventional SG-MOSFET	S_{Vt}	45.839	53.24	55.017	59.936	60.166	66.502	75.868
	S_{IOFF}	1358.282	1430.553	1445.428	1490.462	1500.333	1524.958	1537.601
	S_{SS}	6.0116	7.2199	7.5517	8.923	9.3524	10.924	12.448
	S_{FCR}	11.5616	20.13218	23.52495	45.36716	55.93344	121.5954	268.3308
	I_{ON}/I_{OFF}	46.8831	81.6374	95.3953	183.967	226.814	493.078	1088.1
DM-SGGS-BTFS	S_{Vt}	137.484	158.385	163.811	184.908	191.117	212.456	231.596
	S_{IOFF}	3367.4676	3397.0077	3401.0546	3409.24658	3410.26932	3411.729197	3411.967703
	S_{SS}	6.0618	7.8327	8.3253	10.3935	11.096	13.5831	15.7545
	S_{FCR}	73.07833	213.3475	290.6677	1126.405	1775.47	10886.79	86145.18
	I_{ON}/I_{OFF}	7.3804	21.5466	29.3554	113.759	179.31	1099.49	8700.06
Charged biomolecules		**B_{81}**	**B_{82}**	**B_{83}**	**B_{84}**	**B_{85}**	**B_{86}**	
Conventional SG-MOSFET	S_{Vt}	68.034	79.857	92.847	114.126	182.702	241.128	
	S_{IOFF}	1526.455	1530.443	1534.752	1542.807	1549.301	1549.735	
	S_{SS}	10.9149	10.8865	10.8511	10.7442	10.4423	10.2505	
	S_{FCR}	129.061643	154.413364	196.4094331	406.5429203	4666.701191	28556.10384	
	I_{ON}/I_{OFF}	523.354	626.157	796.454	1648.56	18923.8	115797	
DM-SGGS-BTFS	S_{Vt}	213.277	215.646	218.641	226.528	244.601	256.891	
	S_{IOFF}	3410.74998	3411.803	3414.857	3415.945	3416.998	3420	
	S_{SS}	13.5916	13.6154	13.6438	13.7228	13.7661	13.9795	
	S_{FCR}	11579.8124	13921.65794	17762.22114	36474.01305	441801.9071	1414157.417	
	I_{ON}/I_{OFF}	1169.48	1405.99	1793.86	3683.62	4519.6	142820	
B_{81} : N_f=-1*10^{10}/cm^2	B_{82} : N_f=-4*10^{10}/cm^2	B_{83} : N_f =-8*10^{10}/cm^2		B_{84} : N_f =-2*10^{11}/cm^2		B_{85} : N_f =-6*10^{11}/cm^2	B_{86} : N_f =-1*10^{12}/cm^2	
ϕ_{M1}=5.10 eV [56] , ϕ_{M2}=4.53 eV [56] , ϕ_{M3}=4.10 eV [57]				$\phi_{M(SGM)}$=4.80 eV [75]				

3. Device modelling

3.1 Channel potential

The potential distribution ψ(r,z,T) across the channel satisfies 2D Poisson's equation and is given by,

$$\frac{1}{r}\frac{\partial}{\partial r}\left[r\frac{\partial}{\partial r}\left\{\psi\left(r,z,T\right)\right\}\right] + \frac{\partial^2}{\partial z^2}\psi\left(r,z,T\right) = -\frac{qN_{Cha}}{\varepsilon_{Cha}} \tag{1}$$

where q, $N_{Cha,}$ and ϵ_{Cha} denote the electronic charge, channel doping, and channel permittivity, respectively. The radius of the channel is b, so r=b denotes the surface interface between the channel and oxide layer with potential expressed as $\psi_S(z,T)$ or ψ(b,z,T). ψ(r,z,T) is an implicit function of multiple variables, so its solution follows the parabolic potential profile and can be expressed as

$$\psi\left(r,z,T\right) = M_0\left(z,T\right) + M_1\left(z,T\right)r + M_2\left(z,T\right)r^2 \tag{2}$$

where coefficients M_0, $M_{1,}$ and M_2 (mentioned in Appendix A) be determined using the different boundary conditions.

Various boundary conditions used are :-

(i) $$\psi\left(r,z_1,T\right) = V_{bi,1} = \frac{k_B T}{q}\left\{\ln\left(\frac{N_S N_{Cha_1}}{n_{i(Si)}n_{i(Ge)}}\right)\right\} \tag{3}$$

where, $V_{bi,1}$ denotes the potential barrier at the source-channel junction, $n_{i(Ge)}/n_{i(Si)}$ are the intrinsic concentration of charge carriers in germanium and silicon, k_B denotes the Boltzmann constant (1.38066*10^{-23} J/K), and T denotes the temperature.

(ii) $$\psi\left(r,z_4,T\right) = V_{bi,4} + V_{DS} = \frac{k_B T}{q}\ln\left(\frac{N_{Cha_3}N_D}{n_{i(Si)}^2}\right) + V_{DS} \tag{4}$$

where, $V_{bi,4}$ denotes the potential barrier at the drain-channel junction, and N_D (N_S) denotes the doping of drain (source).

(iii) Electric field is continuous at $z=z_2$ and $z=z_3$,

Accepted Manuscript

$$\left.\frac{\partial\{\psi(r,z,T)\}}{\partial z}\right|_{z=z_2^-} = \left.\frac{\partial\{\psi(r,z,T)\}}{\partial z}\right|_{z=z_2^+}, \quad \left.\frac{\partial\{\psi(r,z,T)\}}{\partial z}\right|_{z=z_3^-} = \left.\frac{\partial\{\psi(r,z,T)\}}{\partial z}\right|_{z=z_3^+} \tag{5}$$

(iv) Electric field inside the bio-tube is zero.

$$\left.\frac{\partial\psi(r,z,T)}{\partial r}\right|_{r=a} = 0 \tag{6}$$

(v) $\psi\left(r,z_2^-,T\right)=V_2$, $\psi\left(r,z_3^-,T\right)=V_3$ (7)

(vi) Electric field is continuous at the Si-SiO_2 interface since electric flux density is the same just above and below the Si-SiO_2 interface in the absence of any interface trap charges.

$$\varepsilon_{Cha}\left.\frac{\partial\{\psi(r,z,T)\}}{\delta r}\right|_{r=b} = C_{ox}\left[V_{GS}-V_{FB}-\psi(r,z,T)\right]\Big|_{r=b}, \quad \left.\frac{\partial\{\psi(r,z,T)\}}{\delta r}\right|_{r=b} = \frac{C_{ox}}{\varepsilon_{Cha}}\left[V_{GS}^{*}-\psi_S(z,T)\right] \tag{8}$$

where, $V_{GS}^{*}=V_{GS}-V_{FB}$ is the effective gate voltage that governs the bending in the energy bands, and C_{OX} / V_{GS} / V_{FB} is the gate oxide capacitance per unit area, gate voltage, and flat-band voltage. $\psi_S(z,T)$ denotes the channel potential at r=b.

vii) $\psi(r,z_2^+)=\psi(r,z_2^-)+V_{bi,2}$, $\psi(r,z_3^+)=\psi(r,z_3^-)+V_{bi,3}$ (9)

viii) $V_{bi,2}=\frac{k_BT}{q}\left\{\ln\left(\frac{N_{Cha_1}N_{Cha_2}}{n_{i(Si)}^2}\right)\right\}$, $V_{bi,3}=\frac{k_BT}{q}\left\{\ln\left(\frac{N_{Cha_2}N_{Cha_3}}{n_{i(Si)}^2}\right)\right\}$ (10)

where, $V_{bi,1}$, $V_{bi,2}$, $V_{bi,3,}$ and $V_{bi,4}$ denote the built-in potential barrier at z_1, z_2, z_3, and z_4, respectively. $N_{Cha,1}$, $N_{Cha,2,}$ and $N_{Cha,3}$ denotes the channel doping in region A, B, and C, respectively. Using the above boundary conditions, it is possible to express the potential in terms of device parameters as given below,

$$\psi(r,z,T)=\psi_S(z,T)+\frac{C_{OX}\left(2ab-b^2\right)\left[V_{GS}^{*}-\psi_S(z,T)\right]}{2\varepsilon_{Cha}(b-a)}-\frac{C_{OX}a\left[V_{GS}^{*}-\psi_S(z,T)\right]}{\varepsilon_{Cha}(b-a)}r+\frac{C_{OX}\left[V_{GS}^{*}-\psi_S(z,T)\right]}{2\varepsilon_{Cha}(b-a)}r^2 \tag{11}$$

For incorporating the effect of temperature, different relations have been formulated as given below [76–80],

$$E_G(T)=E_{G_0}+E_{G_\alpha}\left(\frac{(300)^2}{300+E_{G_\beta}}-\frac{T^2}{T+E_{G_\beta}}\right), \quad \phi_{Sil}(T)=\chi_{Sil}+\frac{E_G(T)}{2}-q\phi_F\left(N_{Cha},T\right) \tag{12a}$$

$$n_i(T)=A_0\left(\frac{T}{300}\right)^{\frac{3}{2}}\exp\left(\frac{-qE_G(T)}{2K_BT}\right), \quad \phi_F\left(N_{Cha},T\right)=\frac{K_BT}{q}\ln\left(\frac{N_{Cha}}{n_i(T)}\right) \tag{12b}$$

$$V_{GS}^{'}(T)=V_{GS}-V_{FB}(T), \quad V_{FB}(T)=\phi_M-\phi_{Sil}(T)-\left(\frac{qN_f}{C}\right) \tag{12c}$$

Values of constant A_0, E_{G0}, $E_{G\alpha}$ and $E_{G\beta}$ for silicon and germanium are $2.2717*10^{19}$, 1.1, $4.73*10^{-4}$, 636 and $0.759144*10^{19}$, 0.66, $4.77*10^{-4}$, 235 respectively [76]. The rest of the symbols has the usual meaning. ϕ_{Sil}/ϕ_M, $\chi_{Sil,}$ and ϕ_F denote the work function of channel/metal, electron affinity, and Fermi potential of channel, respectively. Putting (11) back in (1) yields a differential equation in terms of the surface channel potential $\psi_S(z,T)$,

$$\mu^2\frac{\partial^2\psi_S(z,T)}{\partial z^2}+\theta=\psi_S(z,T) \tag{13}$$

The generalized form of $\psi_S(z,T)$ in three different regions are:-

Region A ($z_1 \le z \le z_2$): $\psi_{S_1}(z,T)=g_1(T)e^{\frac{z}{\mu_1}}+h_1(T)e^{\frac{-z}{\mu_1}}+\sigma_1(T)$ (14)

$$\mu_1=\sqrt{\frac{\varepsilon_{Cha_1}b(b-a)}{C_{OX_1}(2b-a)}}, \quad \sigma_1(T)=V_{GS_1}^{'}(T)+\frac{qN_{Cha_1}b^2}{C_{OX_1}(2b-a)}-\frac{aqN_{Cha_1}b}{C_{OX_1}(2b-a)}, \quad V_{FB_1}(T)=\phi_{M_1}-\phi_{Sil}(T)-\frac{qN_f}{C_2} \tag{15a}$$

$$C_1=\frac{\varepsilon_{OX_{K1}}}{b\ln\left(1+\frac{t_1}{b}\right)}, \quad C_2=\frac{\varepsilon_{bio}}{b\ln\left(1+\frac{t_2}{b}\right)}, \quad C_3=\frac{\varepsilon_{OX_{K2}}}{b\ln\left(1+\frac{t_3}{b}\right)}, \quad C_{OX_1}=\frac{C_1C_2C_3}{C_1C_2+C_2C_3+C_1C_3} \tag{15b}$$

Accepted Manuscript

Region B ($z_2 \le z \le z_3$): $\psi_{S_2}(z,T) = g_2(T)e^{\frac{z}{\mu_2}} + h_2(T)e^{\frac{-z}{\mu_2}} + \sigma_2(T)$ (16)

$$\mu_2 = \sqrt{\frac{\varepsilon_{Cha_2} b(b-a)}{C_{OX_2}(2b-a)}} \ , \ \sigma_2(T) = V'_{GS_2}(T) + \frac{qN_{Cha_2}b^2}{C_{OX_2}(2b-a)} - \frac{aqN_{Cha_2}b}{C_{OX_2}(2b-a)} \quad (17a)$$

$$C_{OX2} = \frac{\varepsilon_{SiO_2}}{b\ln\left(1+\frac{t_{net}}{b}\right)} \ , \ V_{FB_2}(T) = \phi_{M_2} - \phi_{Sil} \ , \ t_{net} = t_1 + \left[\frac{\varepsilon_{K_1}}{\varepsilon_{K_2}}(t_2+t_3)\right] \quad (17b)$$

Region C ($z_3 \le z \le z_4$): $\psi_{S_3}(z,T) = g_3(T)e^{\frac{z}{\mu_3}} + h_3(T)e^{\frac{-z}{\mu_3}} + \sigma_3(T)$ (18)

$$\mu_3 = \sqrt{\frac{\varepsilon_{Cha_3} b(b-a)}{C_{OX_3}(2b-a)}} \ , \ \sigma_3(T) = V'_{GS_3}(T) + \frac{qN_{Cha_3}b^2}{C_{OX_3}(2b-a)} - \frac{aqN_{Cha_3}b}{C_{OX_3}(2b-a)} \quad (19a)$$

$$C_{OX_3} = C_{OX_1} \ , \ V_{FB_3}(T) = \phi_{M_3} - \phi_{Sil} - \frac{qN_f}{C_2} \quad (19b)$$

V_1 (V_4) is the potential at the source (drain) junction, respectively. The coefficients g_1, h_1, g_2 h_2, g_3, and h_3 can be determined using the boundary conditions and can be expressed as,

$$g_1(T) = \frac{0.5}{\sinh\left(\frac{z_1-z_2}{\mu_1}\right)}\left[\sigma_1(T)\left(e^{\frac{-z_1}{\mu_1}} - e^{\frac{-z_2}{\mu_1}}\right) + V_1 e^{\frac{-z_2}{\mu_1}} - V_2 e^{\frac{-z_1}{\mu_1}}\right] \quad (20a)$$

$$h_1(T) = \frac{0.5}{\sinh\left(\frac{z_1-z_2}{\mu_1}\right)}\left[\sigma_1(T)\left(e^{\frac{z_2}{\mu_1}} - e^{\frac{z_1}{\mu_1}}\right) + V_2 e^{\frac{z_1}{\mu_1}} - V_1 e^{\frac{z_2}{\mu_1}}\right] \quad (20b)$$

$$g_2(T) = \frac{0.5}{\sinh\left(\frac{z_2-z_3}{\mu_2}\right)}\left[\sigma_2(T)\left(e^{\frac{-z_2}{\mu_2}} - e^{\frac{-z_3}{\mu_2}}\right) + \left(V_2 + V_{bi,2}\right)e^{\frac{-z_3}{\mu_2}} - V_3 e^{\frac{-z_2}{\mu_2}}\right] \quad (21a)$$

$$h_2(T) = \frac{0.5}{\sinh\left(\frac{z_2-z_3}{\mu_2}\right)}\left[\sigma_2(T)\left(e^{\frac{z_3}{\mu_2}} - e^{\frac{z_2}{\mu_2}}\right) + V_3 e^{\frac{z_2}{\mu_2}} - \left(V_2 + V_{bi,2}\right)e^{\frac{z_3}{\mu_2}}\right] \quad (21b)$$

$$g_3(T) = \frac{0.5}{\sinh\left(\frac{z_3-z_4}{\mu_3}\right)}\left[\sigma_3(T)\left(e^{\frac{-z_3}{\mu_3}} - e^{\frac{-z_4}{\mu_3}}\right) + \left(V_3 + V_{bi,3}\right)e^{\frac{-z_4}{\mu_3}} - V_4 e^{\frac{-z_3}{\mu_3}}\right] \quad (22a)$$

$$h_3(T) = \frac{0.5}{\sinh\left(\frac{z_3-z_4}{\mu_3}\right)}\left[\sigma_3(T)\left(e^{\frac{z_4}{\mu_3}} - e^{\frac{z_3}{\mu_3}}\right) + V_4 e^{\frac{z_3}{\mu_3}} - \left(V_3 + V_{bi,3}\right)e^{\frac{z_4}{\mu_3}}\right] \quad (22b)$$

Boundary conditions at z_2 and z_3 can be used to evaluate V_2 and V_3 due to the continuity of the electric field at z_2 and z_3.

$$\left.\frac{\partial\psi_{S_1}(z,T)}{\partial z}\right|_{z=z_2^-} = \left.\frac{\partial\psi_{S_2}(z,T)}{\partial z}\right|_{z=z_2^+} , \left.\frac{\partial\psi_{S_2}(z,T)}{\partial z}\right|_{z=z_3^-} = \left.\frac{\partial\psi_{S_3}(z,T)}{\partial z}\right|_{z=z_3^+} \quad (23a)$$

$$V_2 = f_4\left(sf_3 - tf_2\right) \ , \ V_3 = f_4\left(sf_2 - tf_1\right) \quad (23b)$$

Values of s, t, and f_1-f_4 are mentioned in Appendix A.

3.2 Threshold voltage

A significant amount of conduction current will flow only after applying a gate voltage greater than the threshold voltage. When the V_{GS} equals V_t, the minimum channel potential is $2\phi_F$. Since drain-to-source voltage decreases while moving towards the source and hence, minimum channel potential occurs in region-A. The threshold voltage is solved by

Accepted Manuscript

substituting the value of minimum channel potential (obtained by differentiating the surface potential), and putting $V_{GS}=V_t$ will yield a quadratic equation in terms of V_t.

$$V_t = \frac{-r_{12} + \sqrt{r_{12}^2 - 4r_{11}r_{13}}}{2r_{11}} \tag{24}$$

Values of r_{11}, r_{12}, and r_{13} are mentioned in Appendix B.

3.3 Drain current and subthreshold swing

The drain current can be modelled for linear region in the three different regions as [28,81,82],

$$I_{Lin,A} = \frac{\pi C_{ox_1}\mu_n (b-a)}{L'_{Cha_1}}\left[(V_{GS}-V_t)(V_2-V_1) - \frac{(V_2-V_1)^2}{2}\right] \tag{25}$$

$$I_{Lin,B} = \frac{\pi C_{ox_2}\mu_n (b-a)}{L'_{Cha_2}}\left[(V_{GS}-V_t)(V_3-V_P) - \frac{(V_3-V_P)^2}{2}\right] \tag{26}$$

$$I_{Lin,C} = \frac{\pi C_{ox_3}\mu_n (b-a)}{L'_{Cha_3}}\left[(V_{GS}-V_t)(V_4-V_Q) - \frac{(V_4-V_Q)^2}{2}\right] \tag{27}$$

Appropriate boundary conditions can be used to calculate V_P and V_Q and can be expressed as,

$$V_P = V_{bi,2} + \psi_{S_1}(r,z_2)\ ,\ V_Q = V_{bi,3} + \psi_{S_2}(r,z_3) \tag{28}$$

The drain current is calculated in the saturation region by replacing V_{DS} with $V_{DS,Sat}$.

$$V_{DS,Sat} = \frac{(V_{GS}-V_t)}{1+\frac{\mu_{efld}(V_{GS}-V_t)}{(L_{Cha_1}+L_{Cha_2}+L_{Cha_3})v_{Sat}}} \quad , \quad \mu_{efld} = \frac{\mu_n}{\{1-\zeta(V_{GS}-V_t)\}\left\{1+\Omega\frac{V_{DS}\mu_n}{(L_{Cha_1}+L_{Cha_2}+L_{Cha_3})v_{Sat}}\right\}} \tag{29}$$

$$\Omega = \left[\frac{V_{DS}\mu_n}{(L_{Cha_1}+L_{Cha_2}+L_{Cha_3})v_{Sat}}\right]\left[1.5+\left\{\frac{V_{DS}\mu_n}{(L_{Cha_1}+L_{Cha_2}+L_{Cha_3})v_{Sat}}\right\}\right]^{-1} \tag{30}$$

where v_{Sat} and μ_{efld} are the saturation velocity of electrons ($v_{Sat}=1*10^7$ cm/s) and maximum low field mobility. ζ is the fitting parameter (0.43), and μ_n is the mobility of electrons. $C_{OX1}/C_{OX2}/C_{OX3}$ is the oxide capacitance per unit area. Subthreshold swing describes the swing in gate voltage obtained when the drain current is changed by a decade and is calculated as given below [83]

$$SS = V_T \ln 10\left\{\left(\frac{\partial\psi(b,z,T)}{\partial V_{GS}}\right)^{-1}\right\}\Bigg|_{Z_{min}} \tag{31}$$

3.4 Sensitivity

Relative change in the device metric (M) has been considered for the investigation of the biosensor and is given below,

$$S_M = |M_{Air} - M_{Bio}| \tag{32}$$

where M_{Bio}/M_{Air} denotes the actual value of metric in the presence/absence of biomolecules and S_M is the sensitivity for the sensing metric 'M'.

4. Results and discussion

When the biomolecules are completely entrapped inside the nanogap cavity, they change the effective gate oxide capacitance, which in turn changes the potential and electric field profile. Figure 3 shows the potential profile across the channel for different biomolecules, while its inset shows the minimum channel potential for different cavity thicknesses. The significant relative change in potential is an excellent qualitative indicator of the improved biosensing ability. A sharp change in the potential can be observed with the increasing dielectric constant of biomolecules. This sharp decrease in the minimum potential across the channel is due to the increased effective gate oxide capacitance. This improves

Accepted Manuscript

coupling and interaction between the charge carriers and the gate. Furthermore, the relative change in potential increases with the increasing K_{bio} [30]. The increase in the thickness of the cavity (t_{CAV}) will increase the number of biomolecules that can be entrapped inside it. This increases the sensitivity of the biosensor. Hence, a significant change in the minimum channel potential can be seen with the changing t_{CAV}.

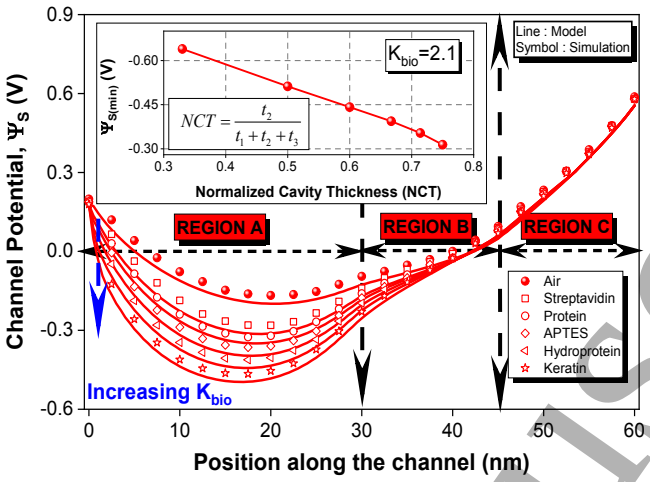


Figure 3. Channel potential profile for different K_{bio} (**[INSET]** shows the minimum channel potential for K_{bio} =2.1 at different normalized cavity thicknesses).

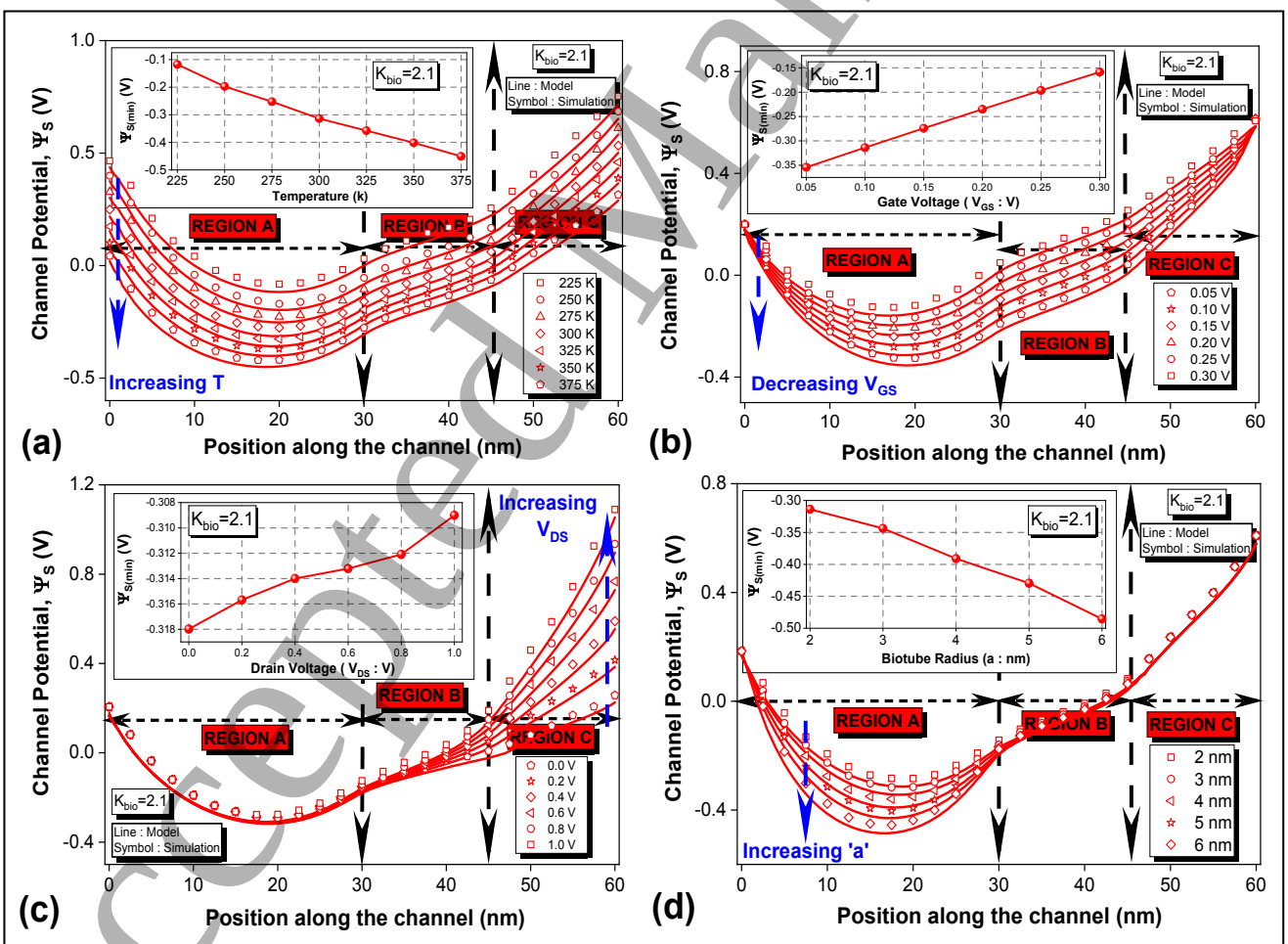


Figure 4. Channel potential profile for K_{bio} =2.1 at **(a)** different temperatures (**[INSET]** shows the minimum channel potential for K_{bio} =2.1at different temperatures), **(b)** different V_{GS} (**[INSET]** shows the minimum channel potential for K_{bio} =2.1at different temperature), **(c)** different V_{DS} (**[INSET]** shows the minimum channel potential for K_{bio} =2.1at different temperature) and **(d)** different biotube inner radius 'a' (**[INSET]** shows the minimum channel potential for K_{bio} =2.1 at different 'a').

Accepted Manuscript

Figure 4(a) shows the potential profile across the channel for K_{bio}=2.1 at different temperatures. Inset shows the minimum channel potential at different temperatures. The number of charge carriers increases with the increasing temperature. Hence, a narrow potential well can be seen at higher temperatures (minimum channel potential keeps on decreasing with the increasing temperature) [84]. Figure 4(b) shows the potential profile across the channel for K_{bio}=2.1 at different gate voltages (V_{GS}). Inset shows the minimum channel potential at different V_{GS} (step size of 50mV). At a fixed drain voltage (V_{DS}), the subthreshold current increases with the increasing V_{GS}. This improvement in the subthreshold conduction is due to the decrease in the potential across the channel, and hence, minimum channel potential also decreases (magnitude) with the increasing gate voltage [84]. It is worth interesting to note that the curvature of the potential curve starts decreasing with increasing V_{GS} because high gate voltage increases the vertical electric field by a significant amount. Figure 4(c) shows the channel potential profile for K_{bio}=2.1 at different drain voltages (V_{DS}). Inset shows the minimum channel potential at different V_{DS}. Changing V_{DS} has a negligible effect on the channel potential variation (especially region A). Increasing V_{DS} will only change the potential barrier at the drain-channel junction; hence, the potential rises near the drain only [84]. It is interesting to note that there is no variation in the potential in region A (near the source) because the potential barrier at the source-channel junction [54] is almost unchanged. Hence, minimum channel potential also infinitesimally increases with the increasing drain voltage. Figure 4(d) shows the potential profile for K_{bio}=2.1 at different inner radius 'a' of the biotube. Inset shows the minimum channel potential at different values of 'a'. With the increasing 'a', the cross-sectional area available for the conducting charge carriers decreases. This increases the threshold voltage of the biosensor due to increasing steepness in the channel potential profile; hence, minimum channel potential also increases (magnitude) with the increasing 'a'.

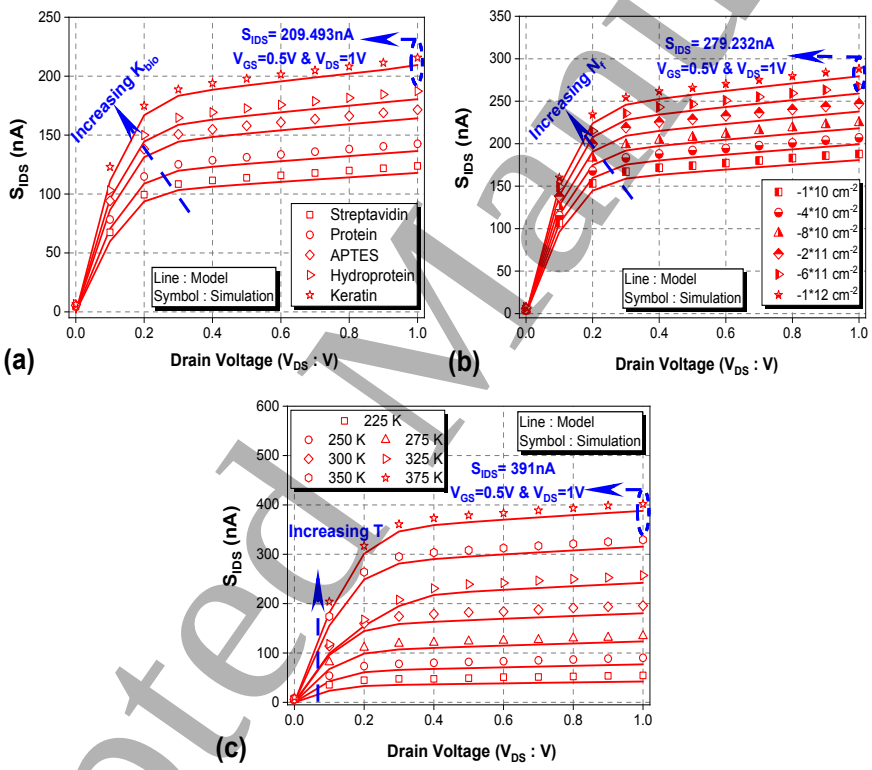


Figure 5. Drain current sensitivity as a function of drain voltage for **(a)** neutral biomolecules, **(b)** charged biomolecules, and **(c)** drain current sensitivity dependency on temperature for K_{bio}=5 at V_{GS}=0.5V.

Figure 5 shows the drain current sensitivity (S_{IDS}) as a function of V_{DS}. Drain current sensitivity increases with the increasing K_{bio}. The biomolecules, when entrapped inside the cavity, will increase the capacitance of the cavity. The increase in the oxide capacitance results in a stronger interaction between the gate and the flow of the charge carriers. This decreases the drain current but the change in drain current increases for the 'unfilled cavity case' [30]. Hence, the drain current sensitivity is high at higher values of K_{bio} which can be seen in figure 5(a). The presence of negatively charged biomolecules results in an increased interaction that accounts for the much stronger gate-to-channel coupling than the neutral biomolecules. Hence, S_{IDS} increases with the increasing charge density of biomolecules, as shown in figure 5(b). The negatively charged biomolecules decrease the channel potential by a factor of qN_f/C (N_f - charge density of biomolecules) in the nanogap cavity [30] that further decreases the drain current, but the change in drain current increases for the 'unfilled cavity case'. Figure 5(c) shows the impact of temperature on drain current sensitivity (S_{IDS}) as

Accepted Manuscript

a function of V_{DS} for K_{bio}=5. As the temperature rises, a large number of charge carriers are generated, which increases the current flowing across the channel [85]. Hence, S_{IDS} increases with increasing temperature. For biosensing applications, operating the device in the ohmic region is advisable, which allows comparatively fast switching and sensing. The maximum S_{IDS} obtained here is roughly 391 nA for K_{bio}=5 at T=375K.

Germanium has a comparatively larger intrinsic charge carrier density than silicon, due to which it offers a large number of intrinsic charge carriers that participates in the conduction process [9] and are also responsible for the biosensing action in DM-SGGS-BTFS. The intrinsic charge carrier density in germanium is approximately ~ 1650 times larger than the silicon at room temperature. When the concentration of charge carriers is high, change in sensing metrics becomes more pronounced as the flow of many charge carriers is affected when different biomolecules localize inside the cavity [10]. Hence, the sensitivity of DM-SGGS-BTFS increases.

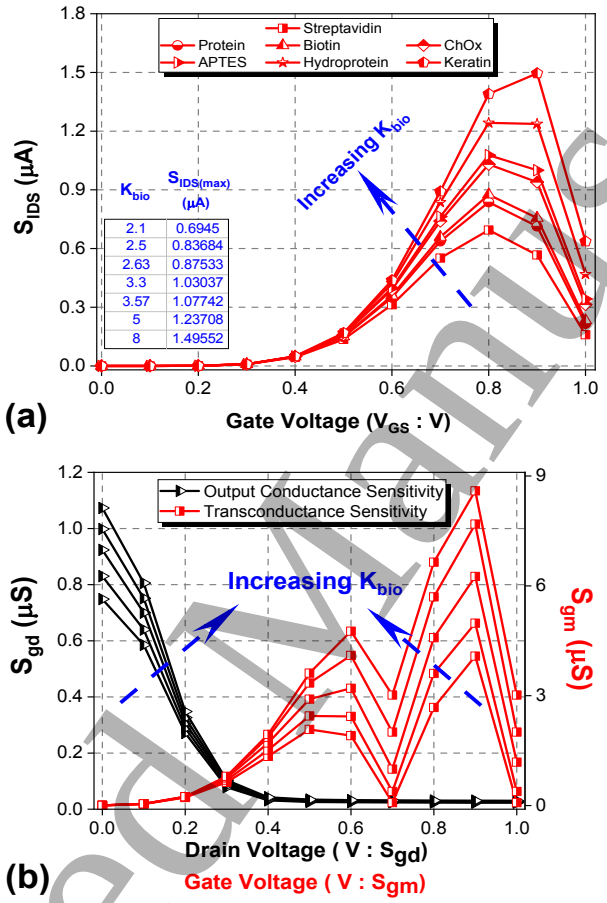


Figure 6. (a) Drain current sensitivity as a function of gate voltage, and **(b)** transconductance (output conductance) sensitivity as a function of gate voltage (drain voltage).

Figure 6(a) shows the drain current sensitivity as a non-linear implicit function of gate voltage. In contrast, figure 6(b) shows the transconductance/output conductance sensitivity (S_{gm}/S_{gd}) for different biomolecules at a fixed value of V_{DS}/V_{GS}. At a lower value of gate voltage, the potential across the channel is nearly constant, but at higher V_{GS}, the potential distribution is not constant (unevenly distributed). It can be inferred that sensitivity is almost constant at lower values of gate voltage, whereas the peak of the drain current sensitivity (S_{IDS}=1.49552μA) is obtained roughly at V_{GS}=0.9V and V_{DS}=0.5V for K_{bio}=8. Further, decreasing the value of V_{DS} will increase the threshold voltage sensitivity and drain current sensitivity. Interestingly, the biosensor is more sensitive to biomolecules at high $V_{GS,}$ and S_{IDS} increases with increasing K_{bio} of the biomolecules. Transconductance is the derivative of drain current obtained from the transfer characteristics of the DM-SGGS-BTFS at a fixed drain voltage value [85]. The slope of the drain current in the I_{DS}-V_{GS} plot is higher at high $V_{GS,}$ and it increases with the increasing $K_{bio,}$ indicating high trans-conductance sensitivity (S_{gm}= 8.5844 μS) is obtained for K_{bio}=8 roughly at V_{GS}=0.9V and V_{DS}=0.5V. We have also used transconductance in determining the sensitivity of the proposed biosensor. Further, the peak transconductance (g_{mp}) can be potentially used as a sensing metric (S_{gmp}) that will increase the reliability of the biosensor. The relative change (compared to the 'unfilled cavity case') in the value of output conductance is high at lower values of drain voltage. This is because the output

Accepted Manuscript

conductance is calculated from the slope of the I_{DS}-V_{DS} curve, and the slope is higher at low V_{DS} (in the ohmic region), which keeps on decreasing with the increasing V_{DS}. This slope tends to zero as the drain voltage increases at a fixed gate voltage value. The maximum output conductance (S_{gd}=1.073 μS) is obtained for K_{bio}=8 at V_{GS}=0.5 V and V_{DS}=0V.

Table 5. Sensitivity variation in DM-SGGS-BTFS at different channel doping.

	STEP-GRADED CHANNEL DOPING (cm^{-3})					BIOMOLECULES: S_{V_t}: Threshold Voltage Sensitivity (mV) $S_{V_t} = \lvert V_t(without_bio) - V_t(with_bio) \rvert$						
	N_S	$N_{Cha,1}$	$N_{Cha,2}$	$N_{Cha,3}$	N_D	B_1	B_2	B_3	B_4	B_5	B_6	B_7
a.		10^{10}	10^{11}	10^{12}		134.629	155.303	160.676	181.603	187.771	209.007	228.094
b.		10^{10}	10^{12}	10^{14}		134.413	155.056	160.423	181.318	187.479	208.688	227.755
c.		10^{10}	10^{14}	10^{18}		137.484	158.385	163.811	184.908	191.117	212.456	231.596
d.		10^{14}	10^{15}	10^{16}		134.135	154.768	160.132	181.025	187.186	208.413	227.523
e.		10^{16}	10^{17}	10^{18}		134.583	155.121	160.458	181.237	187.361	208.441	227.404
f.		10^{14}	10^{16}	10^{18}		137.276	158.17	163.595	184.694	190.906	212.273	231.464
g.		10^{12}	10^{11}	10^{10}		134.742	155.445	160.827	181.796	187.966	209.241	228.371
h.		10^{14}	10^{12}	10^{10}		134.635	155.337	160.72	181.684	187.865	209.159	228.32
i.	10^{19}	10^{18}	10^{14}	10^{10}	10^{19}	46.51	54.367	56.46	64.857	67.423	76.69	85.865
j.		10^{16}	10^{15}	10^{14}		131.771	152.404	157.768	178.661	184.822	206.049	225.159
k.		10^{18}	10^{17}	10^{16}		45.773	53.527	55.593	63.89	66.427	75.597	84.69
l.		10^{18}	10^{16}	10^{14}		46.411	54.251	56.338	64.717	67.278	76.525	85.682
m.		10^{10}	10^{10}	10^{10}		134.851	155.558	160.941	181.901	188.077	209.34	228.448
n.		10^{12}	10^{12}	10^{12}		134.567	155.189	160.582	181.511	187.870	208.002	227.979
o.		10^{14}	10^{14}	10^{14}		134.234	154.876	160.242	181.147	187.312	207.555	227.678
p.		10^{16}	10^{16}	10^{16}		65.983	79.984	83.007	97.459	100.76	111.171	121.873
q.		10^{18}	10^{18}	10^{18}		39.681	46.582	48.432	55.908	58.213	66.611	75.065

a-f: Forward tri-step graded profile (FTSGP) | **g-l:** Reverse tri-step graded profile (RTSGP)

m-q: Non-step graded profile (Mono-doped profile)

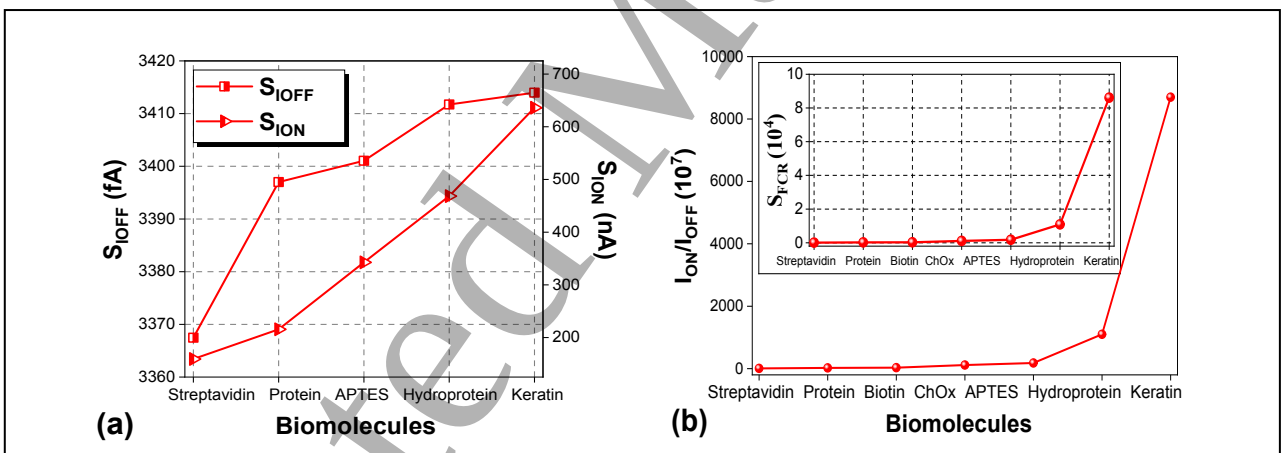


Figure 7. (a) On/off-current sensitivity, and **(b)** I_{ON}/I_{OFF} ratio (**[INSET]** shows the fractional current ratio sensitivity) for different biomolecules.

Using high doping in germanium increases the tunneling probability [8] and facilitates enhanced transportation of charge carriers which is due to the low energy band gap of germanium. As a result, this changes and alters the potential barrier at the source junction, which is bias and electric field dependent). The dependence of different device characteristics (threshold voltage, subthreshold swing, and drain current) on the source-channel potential barrier makes the germanium source biosensors more sensitive to biomolecules.

The graded channel lowers the short channel effects (SCEs) and improves the drive current of the device without scaling it [8],[86]. Using the tri-step graded doping causes a sudden change of potential across the channel, and this sharp change induces a larger electric field than the non-step graded doping [87]. This critical electric field imparts a more vital force on the charge carriers moving from the source to the drain, which improves the carrier transportation mechanism in a graded channel [88]. Using a step graded profile instead of non-step graded profile in channel makes the biosensor more sensitive towards biomolecules. Table 5 shows the sensitivity variation for different channel doping profiles which clearly indicates that the doping of channel must increase gradually while moving from source towards drain to obtain

Accepted Manuscript

high sensitivity. Forward tri step graded profile (FTSGP) yields comparatively high sensitivity as compared to reverse step graded profile (RTSGP), and this sensitivity increases further as the step doping gradient ($|N_{Cha,2} - N_{Cha,1}|$ & $|N_{Cha,3} - N_{Cha,2}|$) is increased. Here, the biosensor has shown maximum sensitivity at $N_{Cha,1}=10^{10}/cm^3$, $N_{Cha,2}=10^{14}/cm^3$ and $N_{Cha,3}=10^{18}/cm^3$ respectively.

Figure 7(a) shows the on-current sensitivity (S_{ION}) and off-current sensitivity (S_{IOFF}) for different biomolecules. The increasing K_{bio} of biomolecules decreases both the on-current and off-current [89], but the relative change increases due to increased gate-to-channel coupling. Hence, S_{ION} [90] and S_{IOFF} increase with the increasing K_{bio}. Figure 7(b) shows the I_{ON}/I_{OFF} ratio, while the inset shows the fractional sensitivity in I_{ON}/I_{OFF} ratio (S_{FCR}) [91]. As both the on-current and off-current decrease, the overall I_{ON}/I_{OFF} ratio increases due to the dominant factor in the denominator (rate of decrease is more in off-current, thus increasing the overall I_{ON}/I_{OFF} ratio with the increasing K_{bio}) [92]. Hence, I_{ON}/I_{OFF} ratio and its fractional sensitivity (S_{FCR}) can also be used as a sensing metric along with threshold voltage and subthreshold swing. I_{ON}/I_{OFF} ratio changes approximately 10^1-10^4 times when the cavity gets filled with biomolecules having K_{bio}=2.1-8. A larger I_{ON}/I_{OFF} ratio variation makes it a perfect sensing metric for biosensing applications. Using multiple sensing metrics to analyze the sensitivity of a biosensor increases its reliability.

Figures 8(a) and 8(b) show the threshold voltage and subthreshold swing sensitivity for different biomolecules. The inset in Fig. 8 shows the variation of threshold voltage and subthreshold swing, respectively. With the increasing K_{bio}, more gate voltage will be required to start the conduction in the device because of the increased capacitance of the nanogap cavity (which leads to high capacitive coupling). This decreases the current flow but increases the threshold voltage. As a result, threshold voltage sensitivity increases due to the increase in the threshold voltage, which can be seen in figure 8(a). Subthreshold swing varies inversely with the gate oxide capacitance, and as a result, subthreshold swing decreases. However, the relative change (with respect to the 'unfilled cavity case') increases with the increasing K_{bio} [90], which is visible in figure 8(b). Threshold voltage has shown a more significant variation than the subthreshold swing for the same biomolecules and hence is a better sensing metric.

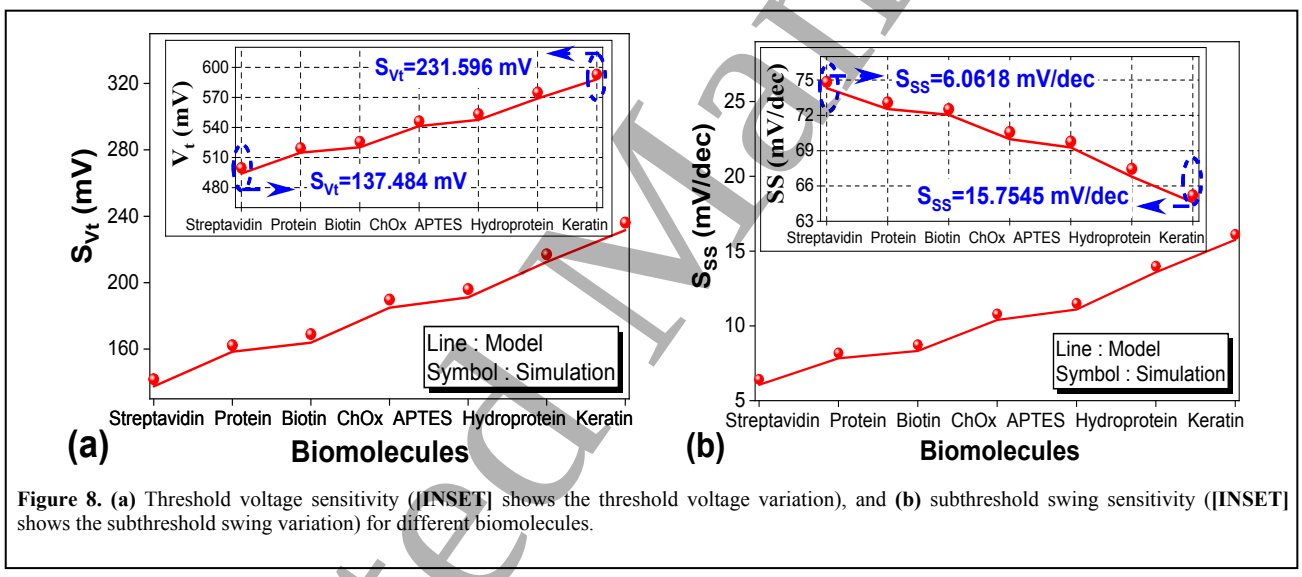


Figure 8. (a) Threshold voltage sensitivity (**[INSET]** shows the threshold voltage variation), and **(b)** subthreshold swing sensitivity (**[INSET]** shows the subthreshold swing variation) for different biomolecules.

Figure 9(a) shows the peak transconductance sensitivity (S_{gmp}), while the inset shows the variation of peak transconductance for different biomolecules. The transconductance is an intrinsic parameter obtained from the I_{DS}-V_{GS} plot. As discussed above, transconductance is high at higher values of V_{GS}, and hence, the peak transconductance sensitivity is also obtained at high V_{GS}. The peak transconductance (g_{mp}=17.3507μS) and peak trans-conductance sensitivity (S_{gmp}=8.5944 μS) is obtained for keratin at V_{GS}=0.9 V and V_{DS}=0.5V. Gate oxide stacking increases the threshold voltage sensitivity by a substantial amount due to the increased gate oxide capacitance (capacitive coupling)[30]. Gate oxide stacking not only improves the sensitivity but also improves the electrostatics integrity of the device. Figure 9(b) shows the effect of the gate stack on threshold voltage sensitivity. A threshold voltage sensitivity improvement of approximately 42mV can be seen for K_{bio}=8 due to gate oxide stacking (SiO_2+HfO_2) which increases the coupling between the gate and the channel when compared to mono gate oxide (SiO_2). We have considered the biotube to be filled with air only, but filling the biotube with other dielectric materials has a negligible effect on the threshold voltage and subthreshold swing. Threshold voltage sensitivity and subthreshold swing sensitivity changes merely 0.00866 V and 0.000537 V/decade when the dielectric constant of the material inside the biotube is changed by 30 times which can be seen in figure 9(c). Therefore, its effect can be neglected (the percentage change is merely below 1% for V_t) in a practical scenario. This minute change is due to the leakage of charge carriers arising due to the non-ideality of

Accepted Manuscript

the dielectric material used in the biotube. More ever, this change can further decrease (negligible) by using an ideal dielectric material inside the biotube.

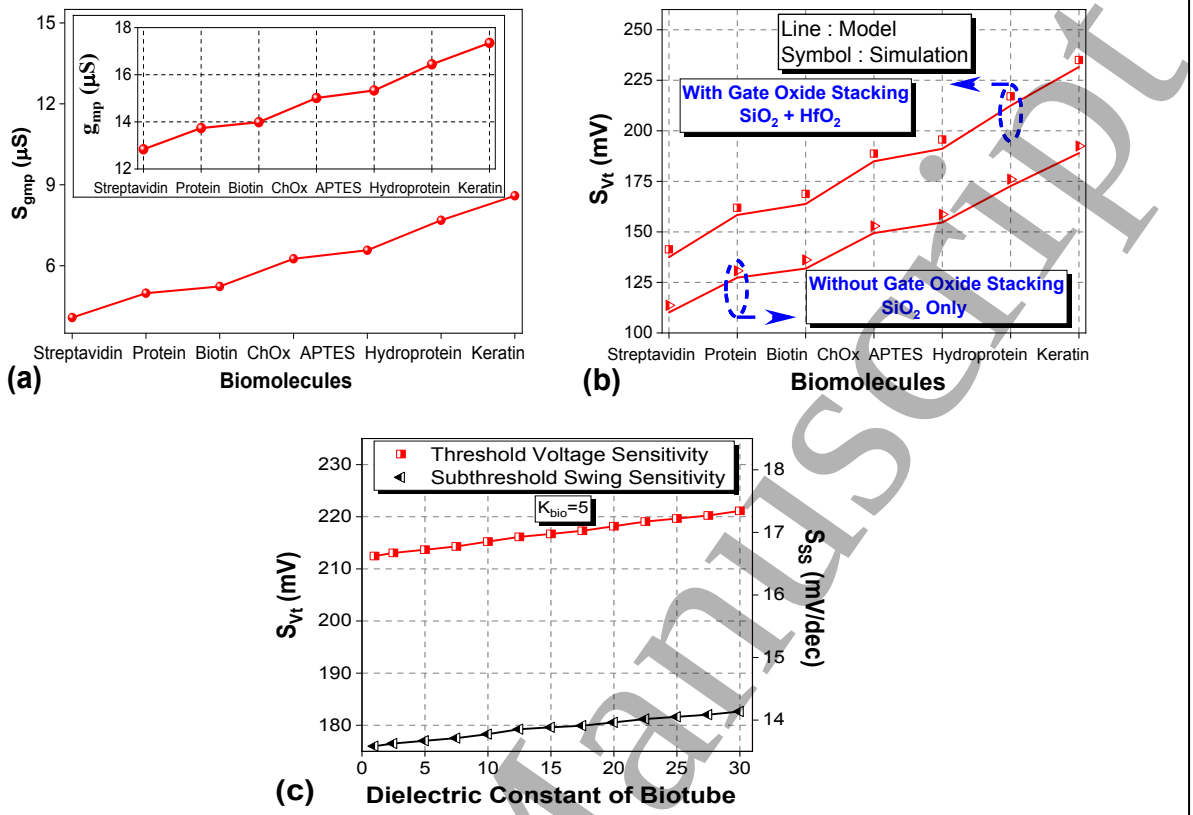


Figure 9. (a) Peak trans-conductance sensitivity (**[INSET]** shows the peak trans-conductance variation), **(b)** effect of gate oxide stacking on threshold voltage sensitivity for different biomolecules, and **(c)** variation of threshold voltage (subthreshold swing) sensitivity due to the dielectric constant of biotube.

Table 6 shows the effect of the fill-in factor, along with the location of biomolecules, on the sensitivity. The fill-in factor implies the fraction of volume within the cavity occupied by the biomolecules. In practical cases, some portion of the cavity might remain unfilled for unavoidable reasons [6]. Hence, studying the impact of the fill-in factor becomes indispensable. It is worth noticing that the sensitivity of a biosensor mainly depends upon two prime factors: i) the location of biomolecules and ii) the fill-in factor of biomolecules. A high fill-in factor implies more biomolecules occupying the cavity which will change the gate oxide capacitance by a larger margin [93]. This changes the electric field and the potential profile across the channel. Hence, the lateral electric field will affect the flow of the charge carriers more strongly. Due to this, the device current changes relatively more [6]. Furthermore, the change in threshold voltage, subthreshold swing, and transconductance will increase if the fill-in factor is high.

Generally, a high fill-in factor of biomolecules results in high sensitivity. However, a high fill-in factor does not always guarantees high sensitivity, which can be seen in different cases. It can be seen that the sensitivity is high when the biomolecules are located near the source. The barrier potential generated across the source junction (between source and channel) mainly affects the device conduction. This potential hill across the source junction is mainly affected by the biomolecules near the source. When the biomolecules are entrapped in the cavity near the source, the potential barrier between the source and the channel is primarily affected. The change in the potential barrier across the source junction is predominantly responsible for a larger change in threshold voltage, transconductance, and subthreshold swing. The charge carriers generated from the source are involved in the biosensing action, and thus the sensitivity is affected by the fill-in factor and location of biomolecules. While the fill-in factor of biomolecules can impact sensitivity, the position of the biomolecules can have a more significant and dominating effect. The fill-in factor in case (3) is higher than in case (4), but the sensitivity relation is opposite to that of the fill-in factor. Similarly, even though case (1) and case (4) have the same fill-in factor but the sensitivity is multiple times larger in case (4) than in case (1), which strengthens the observation that the position of the biomolecules plays an important role in determining sensitivity.

Figure 10 shows the threshold voltage sensitivity plot of DM-SGGS-BTFS at different cavity thickness. Notably, the sensitivity pattern demonstrates a positive correlation with the increasing thickness of the cavity. It is worth noting that, as the cavity thickness increases, the ability of the gate to regulate the flow of charge carriers gradually decreases due to

Accepted Manuscript

the increasing thickness of the gate oxide. Consequently, the rate of sensitivity change becomes lower at higher cavity thicknesses compared to lower cavity thicknesses, as can be seen in figure 10.

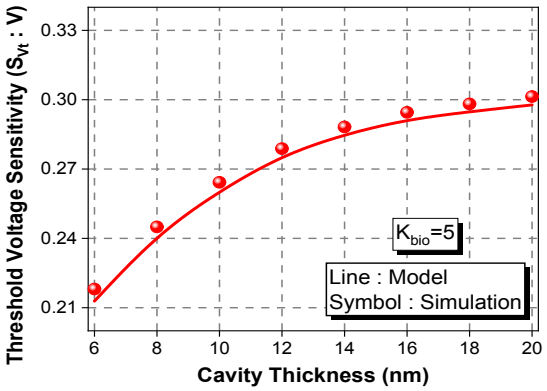
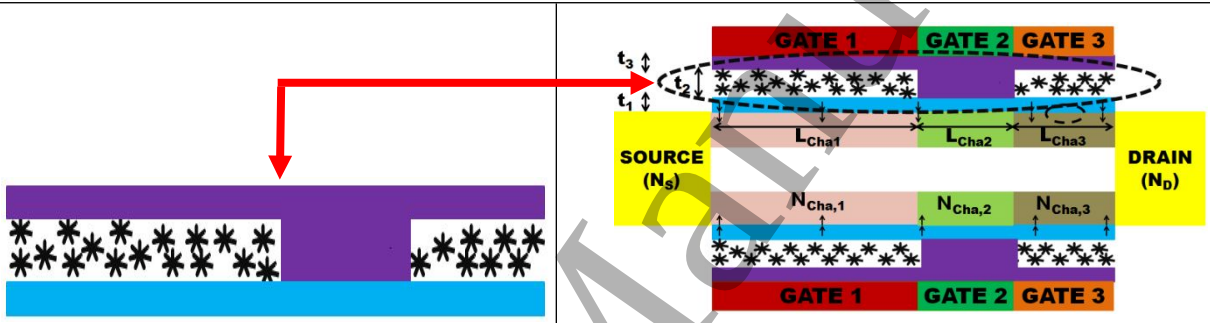


Figure 10. S_{Vt} variation with the cavity thickness in DM-SGGS-BTFS.

Table 6. Effect of fill-in factor and location of different biomolecules on S_{Vt}, S_{SS}, and S_{gmp}.



S. No.	Total cavity on both sides is divided into three equal parts: Cavity towards the source end ($L_{O1}=L_1+L_2$)		Oxide Layer ($L_{O2}=L_3$)	Cavity towards the drain end ($L_{O3}=L_4$)	Fill-in Factor		Threshold Voltage Sensitivity (S_{Vt} : mV) / Subthreshold Swing Sensitivity (S_{SS} : μV/decade) / Peak Transconductance Sensitivity (S_{gmp} : nS) — Increasing K_{bio} → Biomolecules						
	L_1	L_2	L_3	L_4			B_1	B_2	B_3	B_4	B_5	B_6	B_7
1.					0.33	S_{Vt}	4.06	4.521	4.634	5.039	5.143	5.373	5.438
						S_{SS}	0.1446	0.175	0.1832	0.2143	0.2226	0.24	0.2528
						S_{gmp}	260.67	296.33	305.45	340.33	350.36	382.55	403.42
2.					0.33	S_{Vt}	74.035	90.5	95.147	115.04	121.558	147.084	176.305
						S_{SS}	1.325	2.0813	2.3639	3.9885	4.679	7.743	11.1545
						S_{gmp}	1689.18	2150.68	2287.78	2911.28	3129.38	4055.28	5274.18
3.					0.66	S_{Vt}	76.546	92.916	97.511	117.093	123.481	148.408	176.85
						S_{SS}	1.4343	2.2069	2.493	4.1374	4.8319	7.9052	11.3075
						S_{gmp}	1937.68	2425.28	2567.98	3207.38	3427.58	4346.08	5523.88
4.					0.33	S_{Vt}	98.337	118.18	123.64	146.269	153.405	179.88	206.668
						S_{SS}	3.8313	4.8189	5.1036	6.3467	6.7643	8.4622	10.5408
						S_{gmp}	2545.58	3235.28	3438.68	4350.38	4662.28	5926.58	7352.58
5.					0.66	S_{Vt}	100.848	120.645	126.072	148.495	155.541	181.598	207.853
						S_{SS}	3.9554	4.96	5.248	6.5025	6.9224	8.6284	10.7042
						S_{gmp}	2836.98	3572.08	3786.98	4741.58	5064.88	6357.88	7779.08
6.					0.66	S_{Vt}	135.798	156.875	162.359	183.736	190.043	211.754	231.258
						S_{SS}	5.9574	7.7188	8.2146	10.2745	10.9708	13.4767	15.6343
						S_{gmp}	3818.78	4695.58	4940.38	5959.98	6279.38	7417.68	8386.18
7.					1.00	S_{Vt}	137.484	158.385	163.811	184.908	191.117	212.456	231.596
						S_{SS}	6.0618	7.8327	8.3253	10.3935	11.096	13.5831	15.7545
						S_{gmp}	4079.48	4980.78	5230.08	6257.08	6574.48	7686.08	8594.48

Presence of biomolecules	Absence of biomolecules	Oxide layer (HfO_2)
$L_{Cha,1}$=30nm , $L_{Cha,2}$=15nm , $L_{Cha,3}$=15nm	L_{O1}=30nm , L_{O2}=15nm , L_{O3}=15nm	L_{G1}=30nm , L_{G2}=15nm , L_{G3}=15nm
S_{Vt} : Threshold Voltage Sensitivity (mV)	S_{SS} : Subthreshold Swing Sensitivity (mV/decade)	S_{gmp} : Peak Transconductance Sensitivity (nS)

Accepted Manuscript

Figures 11(a) and 11(b) show the effect of temperature on threshold voltage/subthreshold swing sensitivity and I_{ON}/I_{OFF} ratio for K_{bio}=5. As the temperature increases, large numbers of charge carriers are generated due to which conduction starts at a lower value of gate voltage, and hence threshold voltage decreases [92]. But, the relative change in threshold voltage increases with increasing temperature, and hence the threshold voltage sensitivity increases. Due to a substantial increase in current with the rise in temperature, the swing in the drain current also increases, increasing the subthreshold swing sensitivity. This can be seen in figure 11(a). Both the on-current and off-current increase with an increase in temperature [92], but the rate of change of off-current dominates the I_{ON}/I_{OFF} ratio. Hence, I_{ON}/I_{OFF} ratio degrades badly at high temperatures, as seen in figure 11(b). The peak transconductance is directly proportional to the slope of the drain current in the transfer characteristic plot, and the slope increases with an increase in temperature. Therefore, as the temperature increases, the peak transconductance also increases, and the peak transconductance sensitivity also increases accordingly. This trend can be observed in figure 11(c).

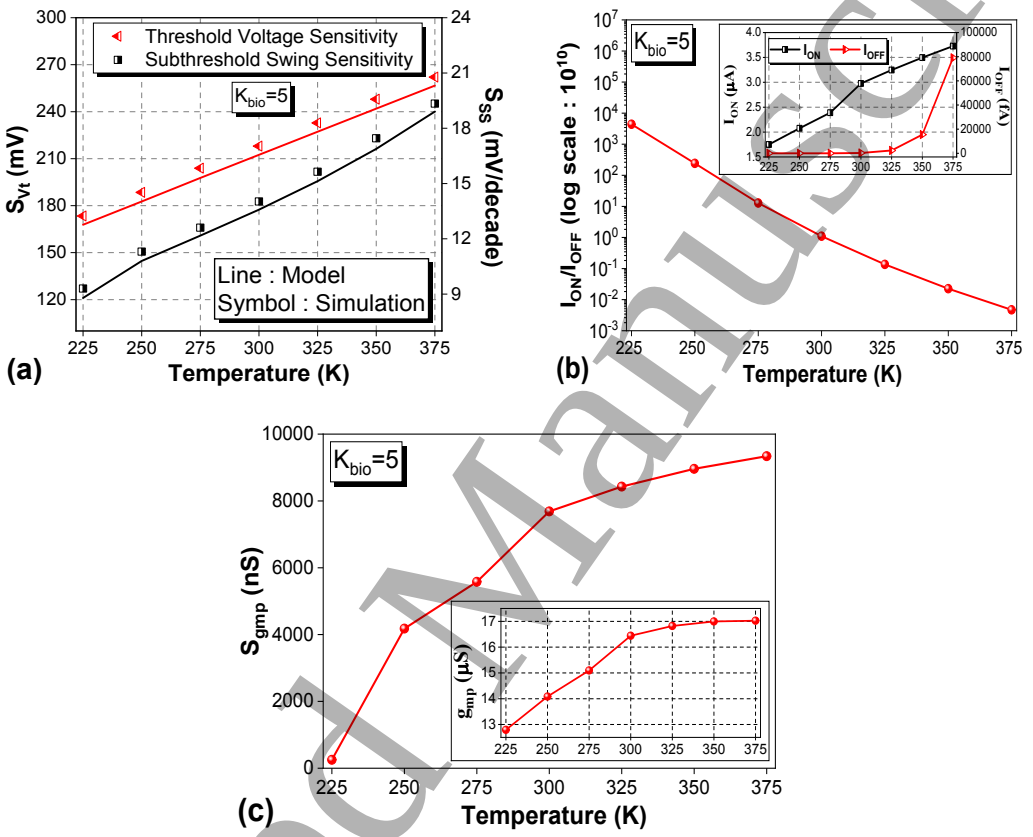


Figure 11. Effect of temperature on **(a)** threshold voltage (subthreshold swing) sensitivity for K_{bio}=5, **(b)** I_{ON}/I_{OFF} ratio for K_{bio}=5 (**[INSET]** shows the variation of on/off current at different temperatures)**, (c)** trans-conductance sensitivity for K_{bio}=5 **([INSET]** shows the peak trans-conductance at different temperatures).

Figures 12(a), 12(b), and 12(c) show the effect of source doping, normalized biotube radius (NBR), and normalized cavity length/thickness on threshold voltage sensitivity. An increase in the source doping will increase the number of charge carriers that brings the nano-device into a conduction state at a low V_{GS} [55,94]. This increases the threshold voltage sensitivity, and the biosensor becomes more sensitive to biomolecules, as seen in figure 12(a). The increasing inner radius of the biotube ('a') will decrease the cross-sectional area of the channel available for the charge carriers, and hence, the conduction current decreases. This will increase the threshold voltage; hence, the relative change in threshold voltage increases with the increasing biotube radius, as seen in figure 12(b). It can be seen that at lower values of 'a' (a=1nm: NBR=0.9), threshold voltage sensitivity is high, but the current ratio is comparatively low. Optimizing the value of 'a' become essential due to the tradeoff between sensitivity and I_{ON}/I_{OFF} ratio. As discussed above, the increasing thickness or length of the cavity will entrap more biomolecules inside it due to its increased volume. Hence, the potential and electric field profile changes relatively more, accounting for the more considerable threshold voltage change. So, threshold voltage sensitivity increases with the increasing cavity length or thickness [30], as seen in figure 12(c). Sensitivity in the label-free biosensors varies directly with the cavity dimensions, i.e., the more is the dimension, the

Accepted Manuscript

more its sensitivity will be [28,30]. Table 7 highlights the sensitivity comparison of DM-SGGS-BTFS with existing biosensors.

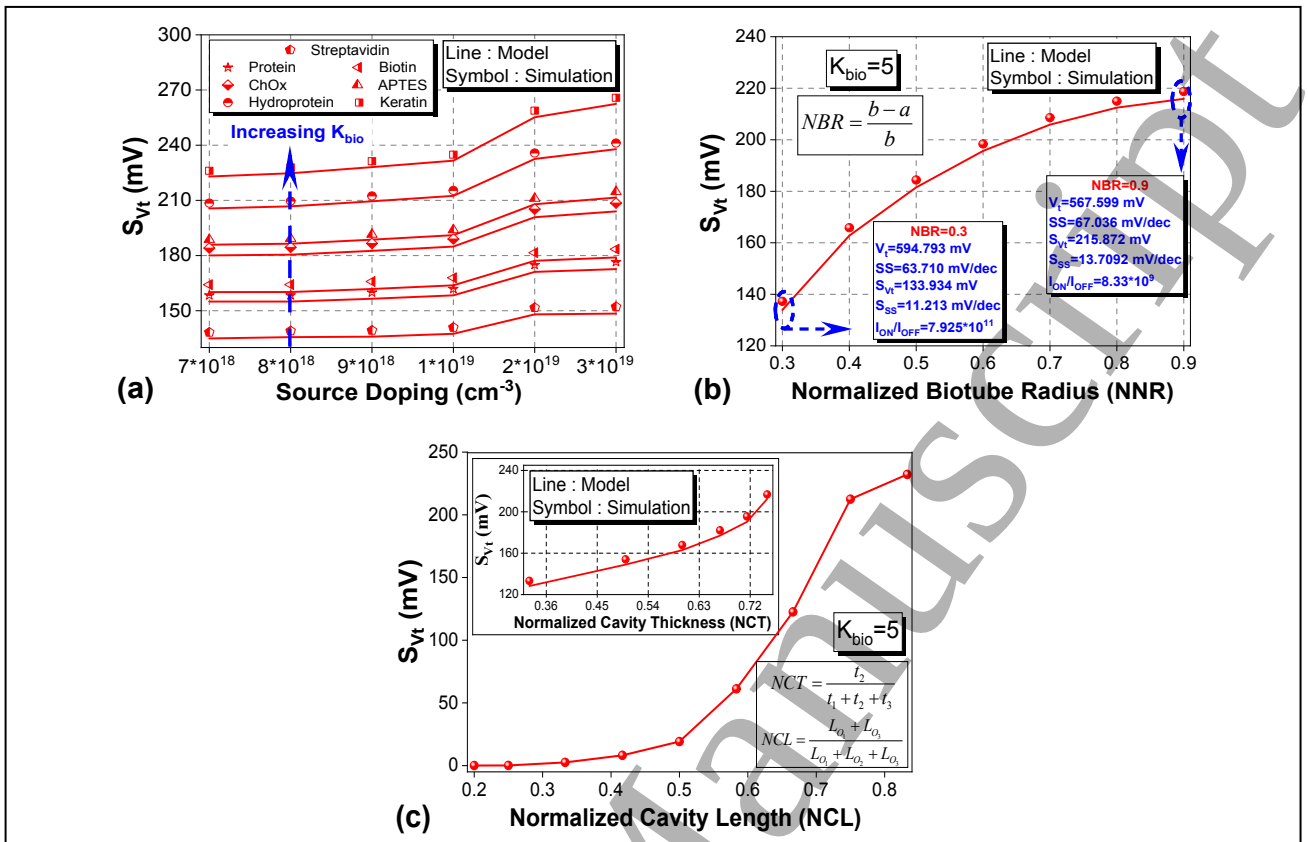


Figure 12. Effect of **(a)** source doping, **(b)** normalized biotube radius, **(c)** normalized cavity length (**[INSET]** shows the variation of threshold voltage sensitivity at different normalized cavity thickness) on threshold voltage sensitivity.

Table 7. Benchmarking the performance of DM-SGGS-BTFS with similar biosensors.

Parameter	Our Work	Reference [95]	Reference [96]	Reference [30]	Reference [97]	Reference [98]	Reference [99]
	DM-SGGS-BTFS	DM-DG-JL-MOSFET	GaN-GME-DE-SNW-FET	JLGSSRG	QG-MC-MOSFET	TGAA-NWFET	TGRC-MOSFET
Gate oxide	SiO_2+HfO_2	SiO_2+TiO_2	Al_2O_3+HfO_2	SiO_2+HfO_2	SiO_2+Si_3N_4	SiO_2+HfO_2	SiO_2
Cavity length (nm)	29	25	15	20	20	10	8
Cavity thickness (nm)	6	9	4	5	9	1	-
Channel length (nm)	60	100	50	50	40	20	40
Type of cavity	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	One-sided	Two-sided
S_{Vt} (V)	0.262	0.227	0.105	0.066	0.161	0.0172	0.0072
K_{bio}	8	10	8	12	12	2.1	8
Advantage	Very high sensitivity	High sensitivity	Low power	Fabrication simplicity	Low dependence of sensitivity on temperature	Low I_{OFF}	Low power
Disadvantage	Fabrication complexity	High threshold voltage	Fabrication complexity	Low sensitivity	Fabrication complexity	Very low sensitivity	Very low sensitivity

Selectivity is one of the most common problems in DM-FET based biosensors (designed without any bioreceptor layer). Hence, this issue has also been addressed here by discussing the merits/demerits of adding an additional bioreceptor layer. For this, the Anti-Iris antibody has been selected as the biomolecule, while the Thiol Linker and Iris Antigen act as the bioreceptor. The respective thicknesses of Thiol linker and Iris Antigen are 2 nm and 4 nm,

Accepted Manuscript

respectively. Although different papers typically consider a thickness of 2 nm for the Anti-Iris antibody, the thickness of the cavity (excluding the bioreceptor thickness) has been increased to 6 nm to accommodate the biomolecule effectively (considering a fill-in factor of 1) [7,9]. The Thiol Linker functions as a silica-binding protein, while the Iris Antigen serves as the bioreceptor for detecting the Anti-Iris (biotarget) layer. These layers possess individual dielectric constants of 2, 4, and 8, respectively. To provide a clear reference, figure 13 visually illustrates the arrangement of the three-layered stack (bioreceptor and biotarget). Within figure 13, the current sensitivity plot (transfer characteristics) for different cases (corresponding to the consecutive stacking of the Thiol Linker, Iris Antigen, and Anti-Iris Antibody) in DM-SGGS-BTFS is presented. In the absence of biomolecules, the drain current remains low. However, as the Thiol Linker, Iris Antigen, and Anti-Iris gradually accumulate within the cavity, the current increases significantly due to the strong electrostatic control of the gate over the channel because of strong capacitive coupling/effect. It can be also seen that threshold voltage sensitivity is higher for the case when no bioreceptor layer is used as compared to the case when bioreceptor layer is used (total oxide thickness, i.e. t_2 is kept same in both the cases). This indicates the existence of a trade-off between sensitivity and selectivity in the proposed device. In non-critical primary sensing applications, the biosensor could be designed without an additional bioreceptor layer, reducing costs and simplifying the manufacturing process (the selectivity/specificity of the biochemical species could be further validated through secondary testing). The inclusion of a bioreceptor layer becomes pertinent when selectivity takes precedence in the sensing application.

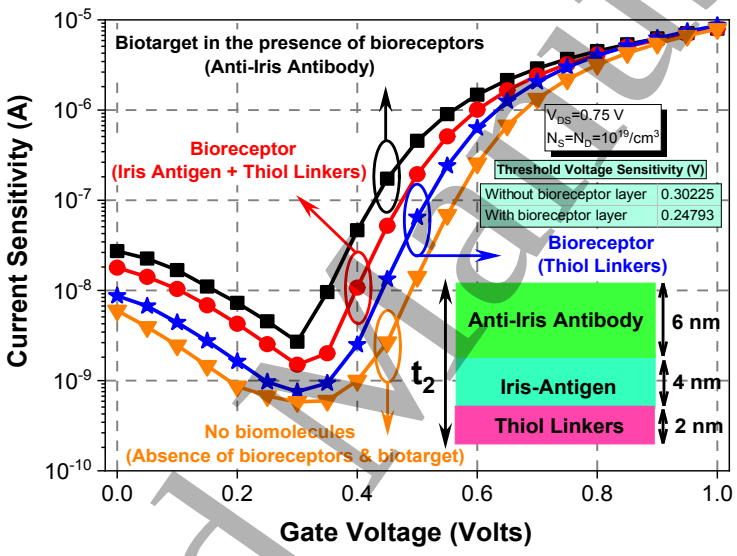


Figure 13. Current sensitivity plot (transfer characteristics) in DM-SGGS-BTFS with successive accumulation of the bioreceptor and biotarget.

4. Conclusion

This paper thoroughly investigates and analyzes the biosensing performance of a novel DM-SGGS-BTFS. The improved biosensing performance is attributed to the use of a germanium source and step-graded biotube, which enhances the biosensing performance and electrostatic integrity of short-channel devices. The study extensively explores the effect of the fill-in factor and the location of biomolecules, demonstrating that the gate controls the flow of charge carriers effectively through capacitive coupling. Notably, the location of biomolecules has a more significant impact on sensitivity than the fill-in factor. Various metrics were used to evaluate the biosensing performance, including subthreshold swing, off-current, threshold voltage, peak trans-conductance, I_{ON}/I_{OFF} ratio, and on-current. The threshold voltage and current sensitivity reached remarkable values of up to 263 mV and 1495.52 nA, respectively, indicating excellent biosensing ability. Furthermore, the study discussed the effect of doping and bias on sensitivity and emphasized the need to optimize different parameters to achieve high sensitivity. For further enhancing the biosensing performance and sensitivity of DM-SGGS-BTFS, future investigations could explore implementing Gaussian/Double-Gaussian doping profile in the nanotube channel, drain engineering techniques using different drain materials, or introducing source/drain pockets. Overall, DM-SGGS-BTFS has the potential to serve as a viable alternative to existing silicon biosensors and can find widespread use in the biomedical industry.

Accepted Manuscript

ACKNOWLEDGEMENT

The authors are grateful to the Director, Maharaja Agrasen Institute of Technology, Delhi for providing the facilities to carry out this research work. One of the authors (Amit Das) would like to thank UGC, Government of India for financial assistance to carry out this research work.

APPENDIX-A

$$M_0(z,T)=\psi_S(z,T)+\frac{C_{OX}\left(2ab-b^2\right)\left[V_{GS}^*-\psi_S(z,T)\right]}{2\varepsilon_{Cha}(b-a)} \quad , \quad M_1(z,T)=-\frac{C_{OX}a\left[V_{GS}^*-\psi_S(z,T)\right]}{\varepsilon_{Cha}(b-a)} \quad , \quad M_2(z,T)=\frac{C_{OX}\left[V_{GS}^*-\psi_S(z,T)\right]}{2\varepsilon_{Cha}(b-a)}$$

$$f_3=2\left[\frac{\beta}{\mu_2}\cosh\left(\frac{z_2-z_3}{\mu_2}\right)+\frac{\gamma}{\mu_3}\cosh\left(\frac{z_3-z_4}{\mu_3}\right)\right] \quad , \quad s=\frac{\alpha}{\mu_1}s_a-\frac{\beta}{\mu_2}s_b \quad , \quad f_1=2\left[\frac{\alpha}{\mu_1}\cosh\left(\frac{z_1-z_2}{\mu_1}\right)+\frac{\beta}{\mu_2}\cosh\left(\frac{z_2-z_3}{\mu_2}\right)\right] \quad , \quad f_2=\frac{2\beta}{\mu_2} \quad , \quad t=\frac{\gamma}{\mu_3}t_a-\frac{\beta}{\mu_2}t_b$$

$$s_a=\left[\sigma_1 e^{\frac{z_2}{\mu_1}}\left(e^{\frac{-z_1}{\mu_1}}-e^{\frac{-z_2}{\mu_1}}\right)-\sigma_1 e^{\frac{-z_2}{\mu_1}}\left(e^{\frac{z_2}{\mu_1}}-e^{\frac{z_1}{\mu_1}}\right)+2V_{bi,1}\right] \quad , \quad s_b=\left[\sigma_2 e^{\frac{z_2}{\mu_2}}\left(e^{\frac{-z_2}{\mu_2}}-e^{\frac{-z_3}{\mu_2}}\right)-\sigma_2 e^{\frac{-z_2}{\mu_2}}\left(e^{\frac{z_3}{\mu_2}}-e^{\frac{z_2}{\mu_2}}\right)+2\cosh\left(\frac{z_2-z_3}{\mu_2}\right)V_{bi,2}\right]$$

$$t_a=\left[\sigma_3 e^{\frac{z_3}{\mu_3}}\left(e^{\frac{-z_3}{\mu_3}}-e^{\frac{-z_4}{\mu_3}}\right)-2V_{DS}-2V_{bi,4}-\sigma_3 e^{\frac{-z_3}{\mu_3}}\left(e^{\frac{z_4}{\mu_3}}-e^{\frac{z_3}{\mu_3}}\right)+2\cosh\left(\frac{z_3-z_4}{\mu_3}\right)V_{bi,3}\right] \quad , \quad t_b=\left[\sigma_2 e^{\frac{z_3}{\mu_2}}\left(e^{\frac{-z_2}{\mu_2}}-e^{\frac{-z_3}{\mu_2}}\right)-\sigma_2 e^{\frac{-z_3}{\mu_2}}\left(e^{\frac{z_3}{\mu_2}}-e^{\frac{z_2}{\mu_2}}\right)+2V_{bi,2}\right] \quad , \quad f_4=\left(f_2^2-f_1f_3\right)^{-1}$$

APPENDIX-B

$$\eta_{11}=4p_{11}-1 \quad , \quad r_{12}=4p_{22}+2\xi_1+4\phi_F \quad , \quad \xi_1=\frac{qN_{Cha_1}ab}{C_{ox_1}(2b-a)}-\frac{qN_{Cha_1}b^2}{C_{ox_1}(2b-a)}+V_{FB_1} \quad , \quad r_{13}=4p_{33}-4\phi_F^2-\xi_1^2-4\phi_F\xi_1 \quad , \quad p_{11}=p_{33}=p_1p_3 \quad , \quad p_{22}=p_1+p_3$$

$$\phi_F=V_T\ln\left(\frac{N_{Cha_1}}{n_{i(Si)}}\right) \quad , \quad \alpha=\frac{1}{2\sinh\left(\frac{z_1-z_2}{\mu_1}\right)} \quad , \quad \xi_3=\frac{qN_{Cha_3}ab}{C_{ox_3}(2b-a)}-\frac{qN_{Cha_3}b^2}{C_{ox_3}(2b-a)}+V_{FB_3} \quad , \quad \xi_2=\frac{qN_{Cha_2}ab}{C_{ox_2}(2b-a)}-\frac{qN_{Cha_2}b^2}{C_{ox_2}(2b-a)}+V_{FB_2}$$

$$m_1=\frac{\alpha}{\mu_1}\left[e^{\frac{z_2}{\mu_1}}\left(e^{\frac{-z_1}{\mu_1}}-e^{\frac{-z_2}{\mu_1}}\right)-e^{\frac{-z_2}{\mu_1}}\left(e^{\frac{z_2}{\mu_1}}-e^{\frac{z_1}{\mu_1}}\right)\right] \quad , \quad m_2=\frac{\beta}{\mu_2}\left[e^{\frac{z_2}{\mu_2}}\left(e^{\frac{-z_2}{\mu_2}}-e^{\frac{-z_3}{\mu_2}}\right)-e^{\frac{-z_2}{\mu_2}}\left(e^{\frac{z_3}{\mu_2}}-e^{\frac{z_2}{\mu_2}}\right)\right] \quad , \quad m_3=\frac{2\beta}{\mu_2}\cosh\left(\frac{z_2-z_3}{\mu_2}\right)V_{bi2}-\frac{2\alpha}{\mu_1}V_{bi1} \quad , \quad p_2=\alpha n_1 \quad , \quad p_4=\alpha n_2 \quad , \quad V_T=\frac{k_BT}{q}$$

$$p_1=\alpha\left[\left(e^{\frac{-z_1}{\mu_1}}-e^{\frac{-z_2}{\mu_1}}\right)-e^{\frac{-z_1}{\mu_1}}f_3f_4(m_1-m_2)-e^{\frac{-z_1}{\mu_1}}f_2f_4(m_4-m_5)\right] \quad , \quad p_3=\alpha\left[\left(e^{\frac{z_2}{\mu_1}}-e^{\frac{z_1}{\mu_1}}\right)+e^{\frac{z_1}{\mu_1}}f_3f_4(m_1-m_2)-e^{\frac{z_1}{\mu_1}}f_2f_4(m_4-m_5)\right] \quad , \quad m_5=\frac{\beta}{\mu_2}\left[e^{\frac{z_3}{\mu_2}}\left(e^{\frac{-z_2}{\mu_2}}-e^{\frac{-z_3}{\mu_2}}\right)-e^{\frac{-z_3}{\mu_2}}\left(e^{\frac{z_3}{\mu_2}}-e^{\frac{z_2}{\mu_2}}\right)\right]$$

$$m_1=\left[\xi_1e^{\frac{-z_2}{\mu_1}}+V_{bi,1}e^{\frac{-z_2}{\mu_1}}-e^{\frac{-z_2}{\mu_1}}e^{\frac{-z_1}{\mu_1}}e_3e_4(-m_1\xi_1+m_2\xi_2-m_3)+e^{\frac{-z_1}{\mu_1}}e_2e_4(-m_4\xi_3+m_5\xi_2-m_6)-\xi_1e^{\frac{-z_1}{\mu_1}}\right] \quad , \quad m_6=\frac{2\beta}{\mu_2}V_{bi_2}-\frac{2\gamma}{\mu_3}\cosh\left(\frac{z_3-z_4}{\mu_3}\right)V_{bi_3}-\frac{2\left(V_{DS}+V_{bi_4}\right)}{\mu_3}V_{bi_2}$$

$$n_2=\left[e^{\frac{z_1}{\mu_1}}f_3f_4(-m_1\xi_1+m_2\xi_2-m_3)-\xi_1\left(e^{\frac{z_2}{\mu_1}}-e^{\frac{z_1}{\mu_1}}\right)-V_{bi,1}e^{\frac{z_2}{\mu_1}}-e^{\frac{z_1}{\mu_1}}f_2f_4(-m_4\xi_3+m_5\xi_2-m_6)\right] \quad , \quad m_4=\frac{\gamma}{\mu_3}\left[e^{\frac{z_3}{\mu_3}}\left(e^{\frac{-z_3}{\mu_3}}-e^{\frac{-z_4}{\mu_3}}\right)-e^{\frac{-z_3}{\mu_3}}\left(e^{\frac{z_4}{\mu_3}}-e^{\frac{z_3}{\mu_3}}\right)\right]$$

$$\xi_2=\frac{qN_{Cha_2}ab}{C_{ox_2}(2b-a)}-\frac{qN_{Cha_2}b^2}{C_{ox_2}(2b-a)}+V_{FB_2} \quad , \quad \xi_3=\frac{qN_{Cha_3}ab}{C_{ox_3}(2b-a)}-\frac{qN_{Cha_3}b^2}{C_{ox_3}(2b-a)}+V_{FB_3}$$

STATEMENTS & DECLARATIONS

CONFLICT OF INTEREST

There are no conflicts of interest amongst the authors.

AUTHOR CONTRIBUTIONS

Amit Das, Sonam Rewari, Binod Kumar Kanaujia, S.S. Deswal and R.S. Gupta have contributed mutually regarding this paper.

AVAILABILITY OF DATA AND MATERIALS

Not Applicable.

FUNDING

Not applicable.

CONSENT TO PARTICIPATE

All the authors have complete consent to participate.

CONSENT FOR PUBLICATION

All the authors have complete consent for publication.

COMPLIANCE WITH ETHICAL STANDARDS

* Disclosure of potential conflicts of interest

There are no conflicts of interest amongst the authors.

* Research involving Human Participants and/or Animals

No animals or human beings were harmed during this research.

*** INFORMED CONSENT**

Accepted Manuscript

All the authors have informed consent.

References

[1] Das A, Kanaujia B K, Nath V, Rewari S and Gupta R S 2020 Impact of Reverse Gate Oxide Stacking on Gate All around Tunnel FET for High Frequency Analog and RF Applications *2020 IEEE 17th India Counc. Int. Conf. INDICON 2020* 1–6

[2] Sharma S, Goel A, Rewari S, Vandana N and Gupta R S 2022 Enhanced Analog Performance and High-Frequency Applications of Dielectric Engineered High-K Schottky Nanowire FET *Silicon*

[3] D. Singh, B. S. Sengar, P. Dwivedi, and V. Garg, "Comparative analysis of gate structure dependent FET-based biosensor," *Mater. Today Commun.*, p. 106301, 2023, doi: https://doi.org/10.1016/j.mtcomm.2023.106301

[4] Im H, Huang X J, Gu B and Choi Y K 2007 A dielectric-modulated field-effect transistor for biosensing *Nat. Nanotechnol.* **2** 430–4

[5] Wadhera T, Kakkar D, Wadhwa G and Raj B 2019 Recent Advances and Progress in Development of the Field Effect Transistor Biosensor: A Review *J. Electron. Mater.* **48** 7635–7646

[6] Das A, Rewari S, Kanaujia B K and Gupta R S 2022 Recent Technological Advancement in Surrounding Gate MOSFET for Biosensing Applications - a Synoptic Study *Silicon* **14** 5133–5143

[7] Dwivedi P and Kranti A 2018 Dielectric Modulated Biosensor Architecture: Tunneling or Accumulation Based Transistor? *IEEE Sens. J.* **18** 3228–35

[8] Lee K W, Choi S J, Ahn J H, Moon D Il, Park T J, Lee S Y and Choi Y K 2010 An underlap field-effect transistor for electrical detection of influenza *Appl. Phys. Lett.* **96** 033703

[9] Tang X, Jonas A M, Nysten B, Demoustier-Champagne S, Blondeau F, Prévot P P, Pampin R, Godfroid E, Iñiguez B, Colinge J P, Raskin J P, Flandre D and Bayot V 2009 Direct protein detection with a nano-interdigitated array gate MOSFET *Biosens. Bioelectron.* **24** 3531–7

[10] Chung I Y, Jang H, Lee J, Moon H, Seo S M and Kim D H 2012 Simulation study on discrete charge effects of SiNW biosensors according to bound target position using a 3D TCAD simulator *Nanotechnology* **23**

[11] Dwivedi P and Singh R 2020 Investigation the impact of the gate work-function and biases on the sensing metrics of TFET based biosensors *Eng. Res. Express* **2** 025043

[12] Sarkar D and Banerjee K 2012 Proposal for tunnel-field-effect-transistor as ultra-sensitive and label-free biosensors *Appl. Phys. Lett.* **100** 143108

[13] Dwivedi P, Singh R, Sengar B S, Kumar A and Garg V 2021 A New Simulation Approach of Transient Response to Enhance the Selectivity and Sensitivity in Tunneling Field Effect Transistor-Based Biosensor *IEEE Sens. J.* **21** 3201–9

[14] Dwivedi P and Kranti A 2018 Overcoming Biomolecule Location-Dependent Sensitivity Degradation Through Point and Line Tunneling in Dielectric Modulated Biosensors *IEEE Sens. J.* **18** 9604–11

[15] Goyal P, Srivastava G, Rewari S and Gupta R S 2020 Controlling Ambipolarity and Rising Ion in TFETs for Enhanced Reliability: A Review *2020 5th IEEE Int. Conf. Recent Adv. Innov. Eng. ICRAIE 2020 - Proceeding* 1–6

[16] Goyal P, Srivastava G, Madan J, Pandey R and Gupta R S 2022 Source material valuation of charge plasma based DG-TFET for RFIC applications *Semicond. Sci. Technol.* **37** 095023

[17] Chen Y N, Fan M L, Hu V P H, Su P and Chuang C Te 2014 Evaluation of stability, performance of ultra-low voltage MOSFET, TFET, and Mixed TFET-MOSFET SRAM cell with write-assist circuits *IEEE J. Emerg. Sel. Top. Circuits Syst.* **4** 389–99

[18] Soni D, Sharma D, Aslam M and Yadav S 2018 Approach for the improvement of sensitivity and sensing speed of TFET-based biosensor by using plasma formation concept *Micro Nano Lett.* **13** 1728–33

[19] Avci U E, Rios R, Kuhn K J and Young I A 2011 Comparison of power and performance for the TFET and MOSFET and considerations for P-TFET *Proc. IEEE Conf. Nanotechnol.* 869–72

[20] Strangio S, Settino F, Palestri P, Lanuzza M, Crupi F, Esseni D and Selmi L 2018 Digital and analog TFET circuits: Design and benchmark *Solid. State. Electron.* **146** 50–65

[21] Reddy N N and Panda D K 2021 A Comprehensive Review on Tunnel Field-Effect Transistor (TFET) Based Biosensors: Recent Advances and Future Prospects on Device Structure and Sensitivity *Silicon* **13** 3085–100

[22] Hwang M T, Heiranian M, Kim Y, You S, Leem J, Taqieddin A, Faramarzi V, Jing Y, Park I, van der Zande A M, Nam S, Aluru N R and Bashir R 2020 Ultrasensitive detection of nucleic acids using deformed graphene channel field effect biosensors *Nat. Commun.* **11** 1543

[23] Sarkar D, Liu W, Xie X, Anselmo A C, Mitragotri S and Banerjee K 2014 MoS2 field-effect transistor for next-generation label-free biosensors *ACS Nano* **8** 3992–4003

[24] Lee J, Jang J, Choi B, Yoon J, Kim J Y, Choi Y K, Myong Kim D, Hwan Kim D and Choi S J 2015 A Highly Responsive Silicon Nanowire/Amplifier MOSFET Hybrid Biosensor *Sci. Rep.* **5** 12286

[25] Jakob M H, Dong B, Gutsch S, Chatelle C, Krishnaraja A, Weber W and Zacharias M 2017 Label-free SnO2 nanowire FET biosensor for protein detection *Nanotechnology* **28** 1–27

[26] Kaisti M, Kerko A, Aarikka E, Saviranta P, Boeva Z, Soukka T and Lehmusvuori A 2017 Real-Time wash-free detection of unlabeled PNA-DNA hybridization using discrete FET sensor *Sci. Rep.* **7** 1–9

[27] Buitrago E, Fagas G, Badia M F B, Georgiev Y M, Berthomé M and Ionescu A M 2013 Junctionless silicon nanowire transistors for the tunable operation of a highly sensitive, low power sensor *Sensors Actuators, B Chem.* **183** 1–10

[28] Pratap Y, Kumar M, Kabra S, Haldar S, Gupta R S and Gupta M 2018 Analytical modeling of gate-all-around junctionless transistor based biosensors for detection of neutral biomolecule species *J. Comput. Electron.* **17** 288–96

[29] Nowbahari A, Roy A and Marchetti L 2020 Junctionless transistors: State-of-the-art *Electron.* **9** 1–22

[30] Chakraborty A and Sarkar A 2017 Analytical modeling and sensitivity analysis of dielectric-modulated junctionless gate stack surrounding gate MOSFET (JLGSSRG) for application as biosensor *J. Comput. Electron.* **16** 556–67

[31] Wang B, Huang W, Chi L, Al-Hashimi M, Marks T J and Facchetti A 2018 High- k Gate Dielectrics for Emerging Flexible and Stretchable Electronics *Chem. Rev.* **118** 5690–754

[32] Lee J C, Cho H J, Katng C S, Rhee S, Kim Y H, Choi R, Kang C Y, Choi C and Abkar M 2003 High-K Dielectrics and MOSFET Characteristics *IEEE Int. Electron Devices Meet.* 4.4.1–4.4.4

[33] Mech B C and Kumar J 2017 Effect of high-k dielectric on the performance of Si, InAs and CNT FET *Micro Nano Lett.* **12** 624–9

[34] Wilk G D, Wallace R M and Anthony J M 2001 High-κ gate dielectrics: Current status and materials properties considerations *J. Appl. Phys.* **89** 5243–75

[35] Kasturi P, Saxena M and Gupta R S 2005 Modeling and simulation of STacked Gate Oxide (STGO) architecture in Silicon-On-

Accepted Manuscript

Nothing (SON) MOSFET *Solid. State. Electron.* **49** 1639–48
[36] Gupta M and Hu V P-H 2021 Influence of Channel Doping on Junctionless and Negative Capacitance Junctionless Transistors *ECS J. Solid State Sci. Technol.* **10** 065009
[37] Fallahnejad M, Vadizadeh M, Salehi A, Kashaniniya A and Razaghian F 2020 Impact of channel doping engineering on the high-frequency noise performance of junctionless In0.3Ga0.7As/GaAs FET: A numerical simulation study *Phys. E Low-Dimensional Syst. Nanostructures* **115** 113715
[38] Kim R, Avci U E and Young I A 2015 Source/drain doping effects and performance analysis of ballistic III-V n-MOSFETs *IEEE J. Electron Devices Soc.* **3** 37–43
[39] Mosfets S D G, Lin H, Member S and Taur Y 2017 Effect of Source – Drain Doping on Subthreshold Characteristics of Short-Channel DG MOSFETs *IEEE Trans. Electron Devices* **64** 4856–60
[40] Curreli M, Zhang R, Ishikawa F N, Chang H K, Cote R J, Zhou C and Thompson M E 2008 Real-time, label-free detection of biological entities using nanowire-based FETs *IEEE Trans. Nanotechnol.* **7** 651–67
[41] Syahir A, Usui K, Tomizaki K, Kajikawa K and Mihara H 2015 Label and Label-Free Detection Techniques for Protein Microarrays *Microarrays* **4** 228–44
[42] Goel A, Rewari S, Verma S and Gupta R S 2018 Dielectric Modulated Triple Metal Gate All around MOSFET (TMGAA)for DNA Bio-Molecule Detection *Proc. Int. Conf. 2018 IEEE Electron Device Kolkata Conf. EDKCON 2018* 337–40
[43] Wu H, Si M, Dong L, Gu J, Zhang J and Ye P D 2015 Germanium nMOSFETs with recessed channel and S/D: Contact, scalability, interface, and drain current exceeding 1 A/mm *IEEE Trans. Electron Devices* **62** 1419–26
[44] Brunco D P, De Jaeger B, Eneman G, Mitard J, Hellings G, Satta A, Terzieva V, Souriau L, Leys F E, Pourtois G, Houssa M, Winderickx G, Vrancken E, Sioncke S, Opsomer K, Nicholas G, Caymax M, Stesmans A, Van Steenbergen J, Mertens P W, Meuris M and Heyns M M 2008 Germanium MOSFET Devices: Advances in Materials Understanding, Process Development, and Electrical Performance *J. Electrochem. Soc.* **155** H552
[45] Saha R, Hirpara Y and Hoque S 2021 Sensitivity Analysis on Dielectric Modulated Ge-Source DMDG TFET Based Label-Free Biosensor *IEEE Trans. Nanotechnol.* **20** 552–60
[46] Zhou J, Zhang C, Liu Q, You J, Zheng X, Cheng X and Jiang T 2020 Controllable all-optical modulation speed in hybrid silicon-germanium devices utilizing the electromagnetically induced transparency effect *Nanophotonics* **9** 2797–807
[47] Lim B S, Arshad M K M, Othman N, Fathil M F M, Fatin M F and Hashim U 2014 The impact of channel doping in junctionless field effect transistor *IEEE Int. Conf. Semicond. Electron. Proceedings, ICSE* 112–4
[48] Zhao H, Zhu F, Chen Y T, Yum J H, Wang Y and Lee J C 2009 Effect of channel doping concentration and thickness on device performance for In0.53 Ga0.47 As metal-oxide-semiconductor transistors with atomic-layer-deposited Al2 O3 dielectrics *Appl. Phys. Lett.* **94** 1–3
[49] Singh K N and Dutta P K 2021 Analytical modeling of underlap graded channel field effect transistor as a label-free biosensor *Superlattices Microstruct.* **155** 106897
[50] Kumar A, Bhusan S and Tiwari P K 2017 A Threshold Voltage Model of Quantum Confinement Effects *IEEE Trans. Nanobioscience* **16** 868–75
[51] Pal A and Sarkar A 2014 Analytical study of Dual Material Surrounding Gate MOSFET to suppress short-channel effects (SCEs) *Eng. Sci. Technol. an Int. J.* **17** 205–12
[52] Singh S, Yadav S and Bhalla S K 2022 An Improved Analytical Modeling and Simulation of Gate Stacked Linearly Graded Work Function Vertical TFET *Silicon* **14** 4647–60
[53] Narang R, Saxena M and Gupta M 2017 Modeling of gate underlap junctionless double gate MOSFET as bio-sensor *Mater. Sci. Semicond. Process.* **71** 240–51
[54] Kumari M, Singh N K, Sahoo M and Rahaman H 2021 Work function optimization for enhancement of sensitivity of dual-material (DM), double-gate (DG), junctionless MOSFET-based biosensor *Appl. Phys. A Mater. Sci. Process.* **127** 1–8
[55] Das A, Kanaujia B K, Deswal S S, Rewari S and Gupta R S 2022 Doping induced threshold voltage and ION/IOFF ratio modulation in surrounding gate MOSFET for analog applications *2022 IEEE International Conference of Electron Devices Society Kolkata Chapter (EDKCON), Kolkata, India* pp 1–6
[56] Das R, Chanda M and Sarkar C K 2018 Analytical Modeling of Charge Plasma-Based Optimized Nanogap Embedded Surrounding Gate MOSFET for Label-Free Biosensing *IEEE Trans. Electron Devices* **65** 5487–93
[57] Kumar M, Haldar S, Gupta M and Gupta R S 2017 Ambipolarity reduction in DMG asymmetric vacuum dielectric Schottky Barrier GAA MOSFET to improve hot carrier reliability *Superlattices Microstruct.* **111** 10–22
[58] Busse S, Scheumann V, Menges B and Mittler S 2002 Sensitivity studies for specific binding reactions using the biotin/streptavidin system by evanescent optical methods *Biosens. Bioelectron.* **17** 704–10
[59] Goel A, Rewari S, Verma S and Gupta R S 2020 Physics-based analytic modeling and simulation of gate-induced drain leakage and linearity assessment in dual-metal junctionless accumulation nano-tube FET (DM-JAM-TFET) *Appl. Phys. A Mater. Sci. Process.* **126** 1–14
[60] Lodhi A, Rajan C, Kumar A, Dip B, Samajdar P and Soni D 2020 Sensitivity and sensing speed analysis of extended nano ‐ cavity and source over electrode in Si / SiGe based TFET biosensor *Appl. Phys. A* **126** 1–8
[61] Kim C H, Jung C, Park H G and Choi Y K 2009 Novel dielectric-modulated field-effect transistor for label-free DNA detection *Biochip J.* **2** 127–34
[62] Ganesh A, Goel K, Mayall J S and Rewari S 2021 Subthreshold Analytical Model of Asymmetric Gate Stack Triple Metal Gate all Around MOSFET (AGSTMGAAFET) for Improved Analog Applications *Silicon* **14** 4063–4073
[63] Li C, Liu F, Han R and Zhuang Y 2021 A Vertically Stacked Nanosheet Gate-All-Around FET for Biosensing Application *IEEE Access* **9** 63602–10
[64] Zhang Y, Han K and Li J 2020 A simulation study of a gate-all-around nanowire transistor with a core-insulator *Micromachines* **11** 1–12
[65] Ye S, Yamabe K and Endoh T 2021 Ultimate vertical gate-all-around metal–oxide–semiconductor field-effect transistor and its three-dimensional integrated circuits *Mater. Sci. Semicond. Process.* **134** 106046
[66] Choi S J, Moon D Il, Kim S, Duarte J P and Choi Y K 2011 Sensitivity of threshold voltage to nanowire width variation in junctionless transistors *IEEE Electron Device Lett.* **32** 125–7
[67] Kumari M, Singh N K and Sahoo M 2022 A detailed investigation of dielectric-modulated dual-gate TMD FET based label-free biosensor via analytical modelling *Sci. Rep.* **12** 1–17

Accepted Manuscript

[68] Narang R, Reddy K V S, Saxena M, Gupta R S and Gupta M 2012 A dielectric-modulated tunnel-FET-based biosensor for label-free detection: Analytical modeling study and sensitivity analysis *IEEE Trans. Electron Devices* **59** 2809–17

[69] Mohanty S S, Mishra S, Mohapatra M and Mishra G P 2022 Hetero Channel Double Gate MOSFET for Label-free Biosensing Application *Silicon* **14** 8109–18

[70] Goel A, Rewari S, Verma S, Deswal S S and Gupta R S 2021 Dielectric Modulated Junctionless Biotube FET (DM-JL-BT-FET) Bio-Sensor *IEEE Sens. J.* **21** 16731–43

[71] Kim S, Ahn J H, Park T J, Lee S Y and Choi Y K 2009 A biomolecular detection method based on charge pumping in a nanogap embedded field-effect-transistor biosensor *Appl. Phys. Lett.* **94** 1–4

[72] Das A, Rewari S, Kanaujia B K, Deswal S S and Gupta R S 2023 Ge/Si interfaced label free nanowire BIOFET for biomolecules detection-analytical analysis *Microelectronics J.* **138** 105832

[73] Software D S 2018 ATLAS User's Manual Device Simulation Software Volume I **I** 318

[74] Sze S M 2017 *VLSI Technology* (McGraw Hill Education)

[75] Hafiz S A, Iltesha, Ehteshamuddin M and Loan S A 2019 Dielectrically Modulated Source-Engineered Charge-Plasma-Based Schottky-FET as a Label-Free Biosensor *IEEE Trans. Electron Devices* **66** 1905–10

[76] Goel A, Rewari S, Verma S and Gupta R S 2019 Temperature-Dependent Gate-Induced Drain Leakages Assessment of Dual-Metal Nanowire Field-Effect Transistor - Analytical Model *IEEE Trans. Electron Devices* **66** 2437–45

[77] Ravindra N M and Srivastava V K 1979 Temperature dependence of the energy gap in semiconductors *J. Phys. Chem. Solids* **40** 791–3

[78] Misiakos K and Tsamakis D 1993 Accurate measurements of the silicon intrinsic carrier density from 78 to 340 K *J. Appl. Phys.* **74** 3293–7

[79] Singh J 2005 *Smart Electronic Materials: Fundamentals and Applications* (Cambridge University Press)

[80] Rewari S, Nath V, Haldar S, Deswal S S and Gupta R S 2018 Gate-Induced Drain Leakage Reduction in Cylindrical Dual-Metal Hetero-Dielectric Gate All Around MOSFET *IEEE Trans. Electron Devices* **65** 3–10

[81] Ajay, Narang R, Saxena M and Gupta M 2017 Modeling and simulation investigation of sensitivity of symmetric split gate junctionless fet for biosensing application *IEEE Sens. J.* **17** 4853–61

[82] Das A, Rewari S, Kanaujia B K, Deswal S S and Gupta R S 2023 Numerical modeling of a dielectric modulated surrounding-triple-gate germanium-source MOSFET (DM-STGGS-MOSFET)-based biosensor J. Comput. Electron

[83] Jung H 2019 Analysis of subthreshold swing in symmetric junctionless double gate MOSFET using high-k gate oxides *Int. J. Electr. Electron. Eng. Telecommun.* **8** 334–9

[84] Agarwal A, Tiwari R, Ranjan S, Pradhan P C and Swain B P 2018 2D Analytical Modeling of Surface Potential for GaAs based Nanowire Gate All Around MOSFET 2D Analytical Modeling of Surface Potential for GaAs based Nanowire Gate All Around MOSFET *IOP Conf. Ser. Sci. Eng.* 012103

[85] Gaspari V, Fobelets K, Velazquez-Perez J E, Ferguson R, Michelakis K, Despotopoulos S and Papavassilliou C 2004 Effect of temperature on the transfer characteristic of a 0.5 µm-gate Si:SiGe depletion-mode n-MODFET *Appl. Surf. Sci.* **224** 390–3

[86] Narendar V and Girdhardas K A 2018 Surface Potential Modeling of Graded-Channel Gate-Stack (GCGS) High-K Dielectric Dual-Material Double-Gate (DMDG) MOSFET and Analog/RF Performance Study *Silicon* **10** 2865–75

[87] Das A, Rewari S, Kanaujia B K, Deswal S S and Gupta R S 2023 Analytical investigation of a triple surrounding gate germanium source metal – oxide – semiconductor field-effect transistor with step graded channel for biosensing applications *Int. J. Numer. Model. Electron. Networks, Devices Fields* 1–25

[88] Pathak V and Saini G 2018 A Graded Channel Dual-Material Gate Junctionless MOSFET for Analog Applications *Procedia Comput. Sci.* 825–31

[89] Shamim Sarker M, Mainul Islam M, Nur Kutubul Alam M and Rafiqul Islam M 2016 Gate dielectric strength dependent performance of CNT MOSFET and CNT TFET: A tight binding study *Results Phys.* **6** 879–83

[90] Jana G, Sen D, Debnath P and Chanda M 2022 Power and delay analysis of dielectric modulated dual cavity Junctionless double gate field effect transistor based label-free biosensor *Comput. Electr. Eng.* **99** 107828

[91] Bind M K and Nigam K 2022 Sensitivity Analysis of Junction Free Electrostatically Doped Tunnel-FET Based Biosensor *Silicon* **14** 7755–7767

[92] Agha F N A K, Hashim Y and Shakib M N 2020 Temperature Impact on the ION/IOFF Ratio of Gate All around Nanowire TFET *2020 IEEE Int. Conf. Semicond. Electron.* 1–4

[93] Rahman E, Shadman A and Khosru Q D M 2017 Effect of biomolecule position and fill in factor on sensitivity of a Dielectric Modulated Double Gate Junctionless MOSFET biosensor *Sens. Bio-Sensing Res.* **13** 49–54

[94] Das A, Rewari S, Kanaujia B K, Deswal S S and Gupta R S 2023 Analytical modeling and doping optimization for enhanced analog performance in a Ge / Si interfaced nanowire MOSFET *Phys. Scr.* **98** 74005

[95] Kumari M, Singh N K, Sahoo M and Rahaman H 2022 2-D Analytical Modeling and Simulation of Dual Material, Double Gate, Gate Stack Engineered, Junctionless MOSFET based Biosensor with Enhanced Sensitivity *Silicon* **14** 4473–84

[96] Sharma S, Nath V, Deswal S S and Gupta R S 2022 Analytical modelling and sensitivity analysis of Gallium Nitride-Gate Material and, dielectric engineered- Schottky nano-wire fet(GaN-GME-DE-SNW-fet) based label-free biosensor *Microelectronics J.* **129** 105599

[97] Maiti S, De A and Sarkar S K 2022 Analytical Modelling of Symmetric Gate Underlap Quadruple Gate Multichannel Junctionless MOSFET Biosensor *Silicon* **14** 6921–32

[98] Getnet M and Chaujar R 2022 Sensitivity Analysis of Biomolecule Nanocavity Immobilization in a Dielectric Modulated Triple-Hybrid Metal Gate-All-Around Junctionless NWFET Biosensor for Detecting Various Diseases *J. Electron. Mater.* **51** 2236–47

[99] Kumar A, Tripathi M M and Chaujar R 2018 Ultralow-power dielectric-modulated nanogap-embedded sub-20-nm TGRC-MOSFET for biosensing applications *J. Comput. Electron.* **17** 1807–15

Accepted Manuscript

Potential of Green Energy: Current Status in Indian Subcontinent

Nikita Jaiswal[1], Ashish Vishwakarma[2] and Kshama Shukla[3]
Department of Civil Engineering, Ajay Kumar Garg Engineering College, 27th Km Stone, Delhi-Meerut Expressway, PO Adhyatmic Nagar, Ghaziabad 201009 UP India
[1]nikkijaiswall@gmail.com, [2]ai2sh2010@gmail.com, [3]kshama.shukla492@gmail.com

Abstract -- **India ranks third in the world in terms of production of green energy. Extensive use of fossil fuels led to environmental issues, like carbon emissions, cleaning the air, mitigating climate change. Alternatives venues like wind, solar, biomass, geothermal, ocean, wave hold promise to provide long-lasting sustainable solutions to these issues. Deployment of green energy sources is needed for energy security and economic development. Currently, green energy accounts for ~37% of India's total energy consumption and its demand is expected to rise by 256 Mtoe by 2040, clocking average annual growth of ~16%. Installed green energy capacity was 159.95 GW in May 2022. This paper identifies the green energy sources available in India and their potential.**

Keywords: Renewable energy, Green energy, Solar energy, Wind energy, Biomass, Geothermal energy, Tidal wave energy

I. INTRODUCTION

GROWING consumption of conventional energy resources gave rise to various problems like emission of greenhouse gases, polluted air and climate change. Such negative environmental impacts require clean energy production for a high quality of life. Rising gap between demand and supply requires increase in green energy with ease in installation procedures and policies.

Green energy is estimated to grow from 41 Mtoe in 2020 to 256 Mtoe in 2040. The energy demand, according to expert discussion, is 1300 TWh of electricity demand and 1345 Mt of coal demand in 2020-21. By March 2014, India had 76.37 GW of installed capacity and it rose to 159.95 GW by May 2022. The installed power capacity in the country was around 408.72 GW as of November 30th, 2022. India has already stated its intention to achieve net-zero emissions by 2070, with renewable energy sources meeting 50% of its energy needs by 2030. This aim is considered sufficient to mitigate climate change. With strong government support, India is ranked third in the world in terms of green energy production capacity.

Adoption of green energy such as solar energy still depends upon various factors such as finances, investments, and consumer behavior. Liming [1] classified the financial systems into several categories, such as government financing, international funding, public stock markets, private sector financing, venture capital, subsidies, etc. Sisodia identified various barriers,investments, subsidies, perceptions, and gaps between green energy generation and implementation [2]. Identified gaps were mostly in investment planning, regulations and finances. Ansari identified 13 barriers that are responsible for delayed solar power plant installation in India [3]. A huge increase in energy demand demands an alternative need for green, workable, inexpensive, and ecologically sound energy sources. The generation of power from green energy sources steadily increased since 2002. Three separate government bodies are working hard to meet target, *i.e.*, MNRE, IREDA and SEC.

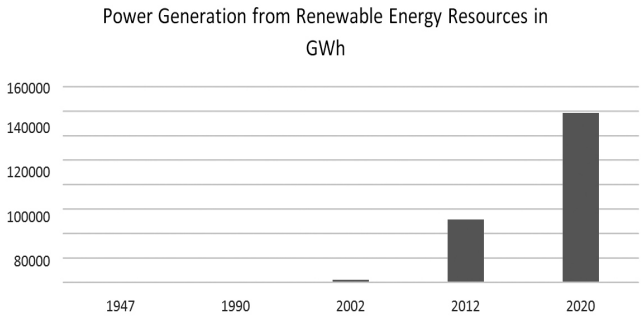


Figure 1. The journey towards renewable energy in India [4]

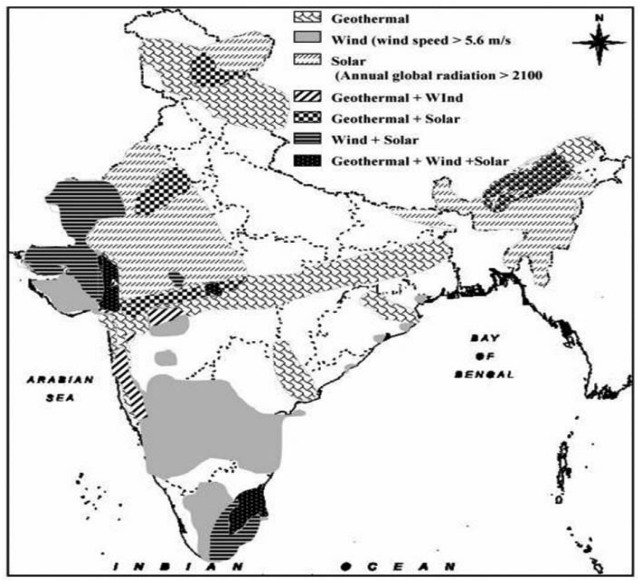


Figure 2. Green resource map India [5].

Commercialized sources of Green energy are as follows.

Geothermal energy: Below the earth's crust, an inexhaustible amount of heat energy exists due to physical processes. Most geothermal resources mapped by the Geological Survey of India, yield estimated potential as about 10600 MW [6]. It is stored in reservoirs beneath the earth's crust in the form of hot water, which is tapped via drilling.

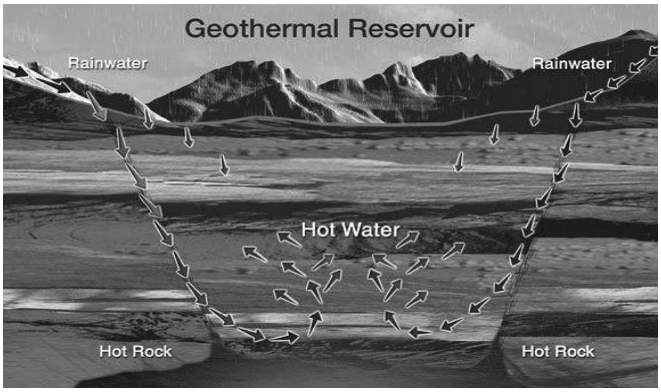


Figure 3. Geothermal reservoir [7].

Wind energy: The commercially exploitable potential of wind energy in India is estimated to be over 200 GW. As of May 2022, the total installed capacity of wind power was 41 GW [9]. The National Institute of Wind Energy (NIWE) installed about 700 wind monitoring stations, out of which 237 have a potential greater than 200 W/m^2. States with great wind energy potential are Rajasthan, Maharashtra, Gujarat, Tamil Nadu, Karnataka, Andhra Pradesh and MP, possessing 293 GW of wind potential at 100 m hub-height and 652 GW at 120 m. The cost of wind power generation is going down day-by-day.

Solar energy: India receives 5000 trillion KWh of solar energy annually. In India, solar incident varies ~ 4-7 kWh/m^2 with bright sunshine. Solar power capacity projected as 749 GW, of which 62 GW was installed as on 30th November, 2022. Solar rooftop market potential is estimated as124 GW in India. Solar rooftop is a distributed system that makes significant contributions to energy independence.

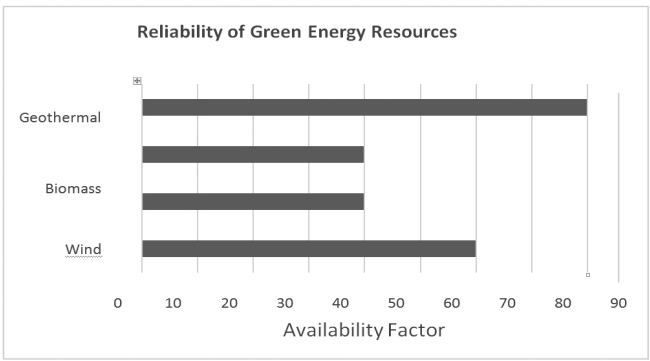


Figure 4. Green sources availability factor [7].

Availability factor of Geothermal sources are more than other kind of Energy sources.

Table 1 -- INDIA'S SOLAR MARK BY 2022 [12]

STATES	SOLAR MARK IN MW
Delhi	2762
Haryana	4142
H.P	776
J & K	1155
Punjab	4772
Rajasthan	5762
U.P	10697
Assam	663
A.P	39
Meghalaya	161
Manipur	105
Mizoram	72
Nagaland	61
Tripura	105

Hydro energy: It is a form of sustainable energy with no carbon emissions. The Himalayan states are capable of producing a large amount of hydroelectric power. MNRE located 6474 potential sites in India: small hydropower capacity is currently projected to be around 20,000 MW. India has so far used 15000 MW of hydropower potential. Worldwide, India ranks fifth in terms of hydroelectric power capacity production. Hydro-electric plants are low-operational cost and require low-maintenance. They provide water and flood control tools.

Biomass energy: The industrial revolution diminished biomass energy, which was the world's dominant energy source before that. India's farming economy relies heavily on bio-mass such as shells, grass, jute, fibre, and coconut shells.The bio-mass can be converted into liquid fuels and electricity. It is estimated that 350 million tons of farm waste are produced each year in India. States with large potential for biomass include UP, Punjab, Karnataka, Gujrat, Bihar and Tamil Nadu. During 2010, India had 3145 MW of installed biomass energy, which rose to 10,228 MW by end-2020. In India, about 56 biogas plants are operating, distributed across Kerala,Maharashtra and Karnataka.

Tidal energy and Wave energy: Tidal energy though not widely used globally, is a good source of green energy. This kind of energy is generated by ocean tides, which are generated by the relative movements among the Sun, Moon, and Earth centrifugal and gravitational forces are responsible for tidal flows or the periodic motion of water. India has a large potential for tidal energy generation due to its 7500 km coastline and 336 islands in the Bay of Bengal and Arabian Sea. The total

available potential of tidal energy is about 40,000 MW. Fluctuations present in tides carry a high power density. This power can be used in a Tidal power plant. Generation of tidal power is dependent on a suitable site, tidal range, and tide speed.

II. CONCLUSION

India's energy requirements are increasing day-by-day, leading to deployment of alternative options like green energy sources to meet the demand. The green sector has significant economic, environmental and social benefits. Strong government support for green energy sector will ensure its greater share in total energy consumption. Literature review reveals spatial variations of different green sources scaled across India.

III. ACKNOWLEDGMENT

The authors gratefully acknowledge Dr. Rakesh Srivastava and Dr. Hemant Ahuja for helping us with the article.

REFERENCES

[1] H. Liming, "Financing rural renewable energy: A comparison between China and India", *Renewable and Sustainable Energy Reviews*, vol. 13, no. 5, pp. 1096-1103, 2009.

[2] G. S. Sisodia, "The Status of Renewable Energy Research on India", *Energy Procedia*, vol. 95, pp. 416-423, 2016.

[3] M. F. Ansari, "Analysis of barriers to implement solar power installations in India using interpretive structural modeling technique", *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 163-174, 2013.

[4] "Renewable Energy in India," 2022. [Online]. Available: https://pib.gov.in/FeaturesDeatils.aspx?NoteId=151141&ModuleId%20=%202.

[5] I. R.Pillai, "Renewable energy in India: Status and potential", *Energy,* vol. 34, no. 8, pp. 970- 980, 2009.

[6] A.Mahesh, "Role of renewable energy investment in India: An alternative to CO2 mitigation", *Renewable and Sustainable Energy Reviews*, vol. 26, pp. 414-424, 2013.

[7] S. Kumar, "Resources and utilization of geothermal energy in India: An ecofriendly approach towards sustainability", *Materials Today: Proceedings*, 2020.

[9] S. Sen, "Renewable energy scenario in India: Opportunities and challenges", *J. African Earth Sciences*, vol. 122, pp. 25-31, 2016.

[10] Harish Puppala, "Identification and analysis of barriers for harnessing geothermal energy in India", *Renewable Energy*, vol. 186, pp. 327-340, 2022.

[11] S. Kumar Kar, "Solar energy market developments in India", *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 121-133, 2016.

[12] M. GOEL, "Solar Roof Top Advancements in India", *Solar Energy*, pp. 149-160, 2022.

[13] M. Irfan, "Prioritizing and overcoming biomass energy barriers: Application of AHP and G- TOPSIS approaches", *Technological Forecasting and Social Change*, vol. 177, 2022.

Nikita Jaiswal is an accomplished Assistant Professor in the Civil Engineering Department at Ajay Kumar Garg Engineering College, Ghaziabad. With a passion for Hydraulics and Flood engineering, she earned her M.Tech degree in this field from Delhi Technological University in 2014. As an Assistant Professor, Nikita brings her extensive knowledge and expertise to the classroom, inspiring and guiding students to excel in their academic pursuits. Her teaching style is engaging and interactive, ensuring that students have a solid understanding of the concepts and practical applications of Civil Engineering.

Nikita's research interests are focused on the development of sustainable and innovative solutions to address hydraulic and flood-related issues. She has published several research papers in reputed international journals.

Ashish Vishwakarma is an Assistant Professor in the Department of Civil Engineering at Ajay Kumar Garg Engineering College, Ghaziabad. He earned his M.Tech degree from IIT BHU Varanasi UP in the domain of Hydraulics and water resources engineering in 2016.

As an Assistant Professor, he brings his experience and expertise in academics to solve physical problems.

His research interests are mainly focused on pipe network,pressure-based fluid mechanics and optimal solutions related to fluid flow. He has published several research papers in reputed international journals.

Ms. Kshama Shukla is an Assistant Professor of civil engineering at AKGEC, Ghaziabad specializing in the field of cognitive struture analysis. She obtained her M.Tech degree in structure analysis from Maharshi Dayanand University, Rohtak, where she conducted research on the mix design of self compacting concrete. She has published several research papers in this field. She is committed to bridging the gap between basic conventional method of construction and new software base technologies of construction. Her current research focus is on 3D concrete printer application and sustainable development.

As an Assistant Professor, she brings her extensive knowledge and expertise to the classroom, inspiring and guiding students to achieve their goals for their bright future. With her interactive and conceptual teaching way, she ensures that students have a solid understanding of the concepts and practical applications of Civil Engineering.

Applied Biochemistry and Biotechnology
https://doi.org/10.1007/s12010-023-04681-w

REVIEW ARTICLE

Potential of *Rhodosporidium toruloides* for Fatty Acids Production Using Lignocellulose Biomass

Sushant Sunder[1] · Anshul Gupta[1,2] · Rashmi Kataria[3] · Rohit Ruhal[3]

Accepted: 16 August 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Microbial lipids are ideal for developing liquid biofuels because of their sustainability and no dependence on food crops. Especially the bioprocess for microbial lipids may be made economical by using sustainable approaches, e.g., lignocellulose-based carbon sources. This demand led to a search for ideal microorganisms with the ability to utilize efficiently biomass into value-added products. *Rhodosporidium toruloides* species belongs to the family of oleaginous (OG) yeast, which aggregates up to 70% of its biomass to produce fatty acids which can be converted to a variety of biofuels. *R. toruloides* is extremely adaptable to different types of feedstocks. Among all feedstock, a lot of effort is going on to develop a bioprocess of fatty acid production from lignocellulose biomass. The lignocellulose biomass is pretreated using harsh conditions of acid, alkali, and other which leads to the generation of a variety of sugars and toxic compounds. Thus, so obtained lignocellulose hydrolysate may have conditions of different pH, variable carbon and nitrogen ratios, and other non-optimum conditions. Accordingly, a detailed investigation is required for molecular level metabolism of *R. toruloides* in response to the hydrolysate for producing desired biochemicals like fatty acids. The present review focuses on numerous elements and obstacles, including metabolism, biofuel production, cultivation parameters, and genetic alteration of mutants in extracting fatty acids from lignocellulosic materials utilizing *Rhodosporidium* spp. This review provides useful information on the research working to develop processes for lignocellulose biomass using oleaginous yeast.

Keywords *Rhodosporidium toruloides* · Biorefinery · Microbial lipids · Lignocellulosic biomass · Oleaginous yeast

✉ Rashmi Kataria
rashmi.kataria@vit.ac.in

✉ Rohit Ruhal
rohit.r@vit.ac.in

[1] Department of Biotechnology, Delhi Technological University, New Delhi, India

[2] Department of Physics, Technische Universität München, Munich, Germany

[3] School of Bio Science and Technology, VIT Vellore, Vellore, Tamil Nadu, India

Published online: 24 August 2023

Introduction

Triacyl glycerides (TAGs) are neutral lipids that are spontaneously synthesized through fatty acid biosynthetic pathways. The application of TAGs is observed in food, animal feed, and feedstocks for many chemical synthesis [1]. Another interesting application of TAGs highlighted by recent research includes its derivatives such as long-chain fatty acid methyl esters (FAMEs) and alkanes as drop-in biofuels [2]. Since the rate of production of drop-in biofuels has its challenges from conventional methods. The methods based on plants and crop oils have a limited capacity and rate of production. Therefore, microbial fatty acid biosynthesis can be used as a tool to overcome these constraints [3].

Oleaginous or oil-producing microbial species are very potent in the conversion of carbohydrates to lipids including TAGs. *Yarrowia* spp., *Cryptococcus* spp., *Trichosporon* spp., *Lipomyces* spp., and *Rhodosporium* spp. are among the intriguing yeast genera that have shown the capacity to collect more than 20% of their dry weight in lipids. However, the *Yarrowia*, *Cryptococcus*, *Trichosporon*, and *Lipomyces* genera require a single-carbon source (typically glucose), in their natural state, to produce high amounts of lipids. The utilization of these strains on substrates that include various forms of carbon sources is severely restricted because of the above limitation. Therefore, in order to enable them to use various feedstocks, it is essential to do extensive research on genetic alterations that may be incorporated [4].

Whereas *R. toruloides*, a non-pathogenic pink-colored basidiomycetous fungus, can accumulate lipids to more than 70% of its dry cell weight and has a remarkable capability for proliferation on a variety of substrates, including glucose, xylose, lignocellulosic hydrolyzate, sludge hydrolyzate, etc. [3, 4]. Additionally, this yeast is an excellent generator of carotenoids and biotechnologically essential enzymes such as epoxide hydrolase and cephalosporin esterase [5]. Furthermore, few reports indicated about its resistance to the inhibitors present in biomass hydrolysate makes it a potential biotechnological yeast strain [6, 7]. There is substantial literature describing the results of lipid production by oleaginous species [8–12], which flourishes our understanding of its genetic and molecular background.

R. toruloides is abundant in nature in several places such as plant leaves, ocean, pine trees pulp, and acid sewage [13–16]. Although in acid sewage they were found in mild acidic environmental conditions. It is categorized as the family Sporidiobolaceae, the order Sporidiobolales, the class Microbotryomycetes, and the phylum Basidiomycota. Mushrooms, smuts, and rusts represent the Basidiomycota phylum's representative fungus that distinguishes *R. toruloides* from several other yeasts used in engineering investigations, including *Saccharomyces cerevisiae*, *Lypomyces starkeyi*, and *Yarrowia lypolytica*. Due to the polyphyletic nature of yeast in the Pucciniomycotina phylum of the Basidiomycota phylum, a new phylogenetic taxonomy advocates renaming *Rhodosporidium* as *Rhodotorula* [17]. *R. toruloides* thrive across various temperatures and pH conditions [3].

The wheat straw, corn straw, rice bran and other plant wastes are sources of lignocellulose biomass. The lignocellulose biomass are carbon sources including oligosaccharides, pentoses and hexoses which can be used for microbial fermentation. Several researchers have developed biorefinery approach to utilize lignocellulose biomass for carbon source. The essential steps involved pretreatment, enzymatic hydrolysis, and fermentation. The main objective remains to remove lignin so that polysaccharides are accessible. But during pretreatment there are chances of generation of toxic components like furfural which can hinder growth of microbes. It is now known that using sugars extracted from lignocellulose may be used for development of economical processes.

Furthermore, these processes are sustainable and environmentally friendly. But we need to identify industrial strains which can rapidly adapt in the hydrolysate with growth inhibitors. Thus, this review focuses on *R. toruloids* ability to utilize lignocellulose waste to produce fatty acids. *R. toruloides* is resistant to biomass-derived inhibitors such as furfural, and hydroxyethyl furfural (HMF) through metabolic modifications [7]. For instance, a study showed how *R. toruloides* strain was adapted in undetoxified hydrolysate to obtain strain with better resistance from the inhibitors like furfural and 5-HMF and improved yield of lipids (sugarcane bagasse hydrolysate was used as a carbon source) [18]. The *R. toruloides* utilizes a variety of carbon sources as energy sources, including oligosaccharides such as cellobiose, maltose, and sucrose; monosaccharides such as pentoses and hexoses; organic acids (OA) such as lactate and acetate; and alcohols such as ethanol, sorbitol, mannitol, and glycerol (Fig. 1) [19, 20]. Several works of literature demonstrate *R. toruloides* may be used to feed on biomass hydrolysates to produce lipids and carotenoids [21–23]. Oleaginous yeasts are often grown in nitrogen-limited conditions to accumulate significant lipids (LPDs) [3]. Additionally, a deficiency of other nutrients such as iron, sulfate, and phosphate facilitate *R. toruloides* lipid synthesis [24].

In the present review, fatty acid production and metabolism in *R. toruloides* is correlated with respect to challenges faced by yeast in culture media made up of lignocellulose biomass hydrolysate. The hydrolysate conditions are very harsh as observed by variety of sugars (pentoses, hexoses), pH, toxic components, and other factors. So main objectives of this review are how pretreatments of lignocellulose waste impact on growth of *R. toruloids*, how composition of hydrolysate can influence production of fatty acids especially C/N ratio, pH, and stress conditions. Furthermore, an effort is done to know if genetic manipulation may help in lipid production and how efficient extraction of lipids is possible.

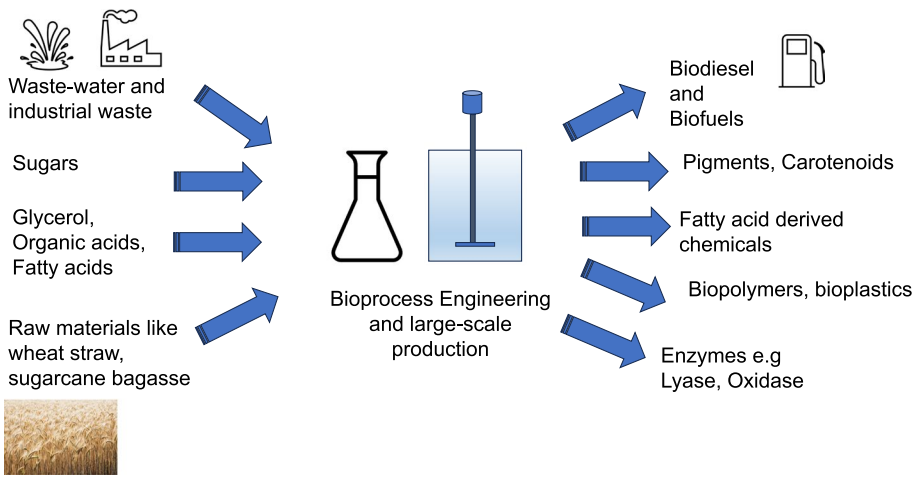


Fig. 1 Schematic representation of value-added produced by *R. toruloides* on uptake from an array of feedstocks like wastewater, fatty acids, organic acids, glycerol, and various sugars

Fatty Acids Production Pathways Known in *R. toruloides*

The production of fatty acids (FAs), their absorption into membrane phospholipids, and storage are dominantly in the form of TAGs, and lastly, the aggregation of stored lipids into lipid droplets as a potential resource is all inherited from the biosynthetic pathway [4]. The major portion of these droplets is utilized as fatty acids synthesized from de novo pathways and stored in the cytoplasm. In the past, lipid droplets were believed to be intracellular fat reservoirs since they retain more than 90% (w/w) of the neutral lipids including steryl esters and TAGs in their whole cell volume [25]. TAG to steryl esters ratios differ across species and are also affected by growth conditions [26]. In *R. toruloides,* FAs are synthesized through de novo pathways, external FAs uptake, or lipid breakdown. De novo fatty acid production occurs in cytoplasm and mitochondria, where it is produced as carbon chains with a stretch of C-18 that are then lengthened and desaturated in the endoplasmic reticulum (ER). Acetyl-CoA (AC), the active precursor for the de novo FAs pathway, originates from citrate via mitochondrial Tricarboxylic acid cycles (TCA cycles). Supplementary precursor citrate is available in oleaginous organisms, decomposed by ATP-citrate lyase (ACL), which affects lipid deposition.

The action of the TCA cycle enzyme iso-citrate dehydrogenase in *R. toruloides* is dependent on intracellular adenosine-monophosphate (AMP) concentrations, and the latter is an allosteric stimulator of this enzyme (Fig. 2). The enhanced action of mitochondrial AMP deaminase decomposes AMP when nutrient becomes scarce. A minimal amount of AMP inhibits iso-citrate dehydrogenase activity. The MPDH circuit is named after the action of mitochondrial pyruvate dehydrogenase (MPDH), which synthesis citrate that is translocated to the cytoplasm and reduced to acetyl coenzyme (Fig. 2) [27]. As a result, the MPDH pathway provides cytoplasmic AC in *R. toruloides*.

In FAs biosynthetic pathways, the acetyl-CoA carboxylase (ACC) protease converts acetyl-coenzyme to C3 malonyl-CoA in a condensation reaction. In *R. toruloides*, the cytoplasmic FA composition is introduced by ACC coded by the ACC1 allele, whereas the mitochondrial FA composition is undertaken by mitochondrial ACC [28]. The resultant

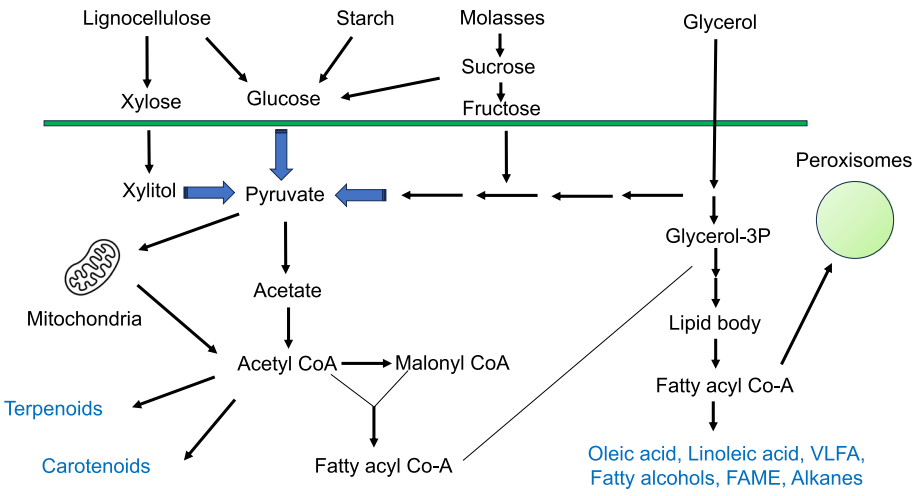


Fig. 2 The figure represents metabolic pathways linked with fatty acid biosynthesis in *R. toruloides*

malonyl-CoA is subsequently extended to palmitoyl (C16)-CoA by a sequence of stages regulated by a cytoplasmic fatty acid synthase (FAS) complex [29]. The acyl chain needs a constant supply of NADPH for this fatty acid synthase complex, which is provided by the malic enzyme. The cytoplasmic upregulation of malic enzyme in oleaginous microbes results in a 2.5-fold rise in lipid synthesis, any mutation in malic enzyme reduces FA accumulation [30]. Throughout the ER, the fatty acyl chain is further elongated, hydroxylated, and desaturated. Whereas various elongates carry out the earlier, desaturation is usually carried out by Δ15-desaturases found in oleaginous yeasts, for the extraction of α-linolenic acid from linoleic acid. Different transcription factors (TFs) control FA desaturation at the transcriptional level [29].

Different Pretreatment Methods Applied on Lignocellulose Biomass Before Using for Fatty Acid Production by *R. toruloides*

Because of its affordability, availability, and sustainability, lignocellulosic biomass has been extensively recognized as a possible substrate for the manufacture of biofuels and biochemicals. Lignocellulosic biomass is made up mostly of cellulose, hemicellulose, and lignin, which are linked together via covalent bonds, intermolecular linkages, and Van der Waals interactions to form an integrated complex that is resistant to hydrolysis and immiscible. Cellulose (40–50% of total dry matter) is a glucose polymer united with β-1,4 glycosidic linkages, with cellobiose, a glucose dimer, as the fundamental building component [31]. Hemicellulose is a highly branched heteropolymer consisting of multiple carbohydrates that accounts for 20–40% of total dry matter (C5 and C6). Depending on the plant species, the specific carbohydrates in hemicellulose might vary. Hemicellulose hydrolysis, on the other hand, may provide a combination of glucose, galactose, mannose, arabinose, xylose, and rhamnose, with lignin accounting for 20–25% of the total dry matter of the substrate [32]. Due to these advantages, lignocellulosic biomass is considered a vital source of lipids through oleaginous microbes. Several trials have been conducted (as described in Table 1) with lignocellulosic biomass with *R. toruloides* fermentation process with good lipid yields. For instance, SànchezNogué et al., in their study, successfully extracted ~ 60% lipid from corn stover hydrolysate [7]. Similarly in 2018, Slininger in his work abstracted the lipid titer of ~ 26.4 g/L from lignocellulosic rich switchgrass hydrolysate [33].

Numerous pre-treatment methods were developed over the last few centuries to resolve biomass intransigence by interrupting the physical boundaries of the cell membrane, lowering cellulose crystallinity, and expelling lignin and hemicellulose to facilitate the enzymatic hydrolysis of lignocellulosic biomass into simple sugars [36]. Pre-treatment methods are broadly divided into four primary groups: physical, chemical, biological, and combination pre-treatment methods. Chemical treatment is defined as the application of organic or inorganic substances that disturb the formation of lignocellulose by interacting with the intra-polymer or interpolymer linkages of lignin, hemicellulose, and cellulose [37]. Acids, alkalis, ionic liquids (ILs), oxidizing agents, and organosolv pretreatment are some of the most often utilized methods. One of the more successful techniques of dissolving the hemicelluloses and making cellulose quite accessible is to pretreat lignocelluloses with acids. The key drawbacks of this methodology are the corrosive environment of reactions and possible formation of inhibitors hydroxymethylofurfural (HMF), furfural, and acetic acid of further processing [38], loss of products because of by-products such as sugar degradation products (e.g., hydroxymethylfurfural or HMF and levulinic acid), hemicellulose

Table 1 Lipid accumulation through different lignocellulosic biomass and *R. toruloides* strains

S. no	Microbial Strain	Substrate	Sugar (g/L)	Lipid titer (g/L)	Lipid (%)	Reactor type	Reference
1	*R. toruloides* DMS—4444	Corn stover hydrolysate	100	23.3 ± 1.8	60.8 ± 1.1	Batch	[7]
2	*R. toruloides* DRRL Y-1091	Switchgrass hydrolysate	215	26.2 ± 1.4	61.5	Flask	[33]
3	*R. toruloides* AS 2.1389	Sugarcane Bagasse hydrolysate	20	12.3 ± 0.5	52.5 ± 1.0	Batch	[6]
4	*R. toruloides* CCT 0783	Sweet Sorghum	140	13.77	33.1	Flask	[34]
5	*R. toruloides* IFO 0880	IL pretreated corn stover hydrolysate	73.8	-	27	Batch	[35]

degradation products(e.g., ferulic acid, glucuronic acid, and p-coumaric acid) and lignin breakdown products (e.g., syringaldehyde and syringic acid) [39]. Biological pretreatment can be executed by utilizing fungi, that decompose lignin (white-rot fungi), cellulose (soft-rot fungi), or hemicelluloses and cellulose (brown-rot fungi). This approach has many benefits, including minimal equipment, efficacious cellulose and hemicellulose degradation, and minimal energy demands [39]. Particularly, enzymatic hydrolysis has been considered a viable approach of hydrolysis due to lower energy utilization, lower production of inhibitors under mild condition, and no discharge of toxic compounds. Apart from this, the yield of sugar obtained from this methodology is very low because the cellulose is parallelly consumed by the enzymatic microbes during the growth phase. Enzymatic hydrolysis generally takes 30–60 days long reaction for the decomposition resulting a major limitation for industrial utilization [40]. Another negative is the waste of material that might have been used for anything else.

Hydrolysis is one of the crucial aspects to extract sugars from pretreated lignocellulosic biomass [41]. Because of lignocellulosic resistance to hydrolysis, it must be transformed effectively and cost effective into its three polymeric components as cellulose, hemicellulose, and lignin, to maximize its significance for biofuel production. During the dilute acid treatment, hemicellulose and cellulose are hydrolyzed to produce fermentable sugars. In principle, the breakdown of hemicellulose results in the production of organic substances such as formic, acetic acid, and carbohydrates (glucose, xylose, mannose, and galactose). Therefore, to effectively transform carbon polymers into simple sugars, pretreatment of biomass materials is often performed before saccharification [42]. Inappropriate management of the pretreatment technique has a vast impact on the process and may result in the formation of substituents as an inhibitory influence during the extraction of biomass. The application of chemical and enzymatic procedures is often employed in the hydrolysis of the cellulose part of the biomass. The presence of acid and heat compels the polysaccharides to undergo chemical breakdown. It happens quickly and without the need for pretreatment of the feedstock. The downsides of this technique include the production of furfurals, poor carbohydrate extraction, and the lower economic viability application. The only way to make the procedure cost-effective is to design efficient methods for recovering acid. Additionally, enzymatic hydrolysis is carried out at moderate heat which varies from 45 to 50 °C, pressure gradient, and a mild pH which results in large-scale production of sugars with less by-product generation and prevents equipment failure. High levels of catalysts are thus necessary for the said decomposition of cellulose, and their manufacture requires an understanding of effective microbial strains and practical, cost-effective methods.

The Impact of Properties of Pretreated Lignocellulose Biomass Hydrolysate for Fatty Acid Production in *R. toruloides*

Fermentation is largely influenced by the environmental variables and cultivation techniques, including the C6-C5 fermentation sugar presence, the quantity of ammonia that is accessible, the carbon to nitrogen ratio (C/N), ionic strength, and biological oxygen demand, along with other things. Anti-proliferative substances found in lignocellulosic materials extracts, however, have an adverse effect on these variables because they interfere with the absorption and assimilation of various carbon sources [18]. In the present part the factors in media like carbon source, pH, nitrogen, and oxygen source are discussed with respect to known factors for fatty acid production in *R. toruloides*.

Ability of *R. toruloides* to Consume Diverse Sugars in Hydrolysate

The pretreatment led to the release of different type of sugars like pentoses, hexoses, and oligosaccharide. The various studies indicated use of different sugars on lipid accumulation. When cells are stressed with sufficient carbon sources (especially glucose) and a critical nutrient deficiency in the growth media, the genus *Rhodosporidium* efficiently utilizes sugars to form FAs [4]. This form of yeast alters the carbon flow (by halting the synthesis of nucleic acids and proteins) to synthesize lipids [43]. This is highly dependent on the environment and culture techniques, including the carbon supply, the amount of accessible nitrogen, O_2, pH, and the carbon–nitrogen (C/N) ratio [4]. Nonetheless, these characteristics are indirectly impacted by inhibitory chemicals found in lignocellulosic biomass hydrolysates since they affect the absorption and assimilation of the hydrolysate's various carbon sources. *Rhodosporidium toruloides* has an excellent capacity to feed on hexose and pentose carbohydrates that includes arabinose, xylose, and glucose [4]. However, fatty acids production varies with type of sugar are used as carbon source for instance, the lipid yield was 75% when glucose was used, 45% with xylose, and 19% with arabinose [44]. Even combination of carbon sources did not result in enhanced yield (30% of lipid accumulation) [45]. In this regard, the optimal utilization rates for glucose and xylose were calculated using a C/N ratio of 100 for lipid accumulation and 65 for biomass formation [23]. The ratios mentioned above result in a substantial change in the parameters and formulation of the fermentation technique to maximize sugar utilization [4, 45]. This adequate ratio of C/N being greater than 65 affects the repression of glucose during the high concentration in media. This results in the osmotic shock of Krebs cycle activity due to intracellular increase of citric acid. As a result, nitrogen limitation is observed, and microbes uses carbon source to formulate lipid as a protective mechanism [4].

The Variable C/N Ratio in Hydrolysate Impact Lipid Production in *R. toruloides*

This parameter is critical for the effective utilization of carbon sources for FAs conversion since it impacts biomass synthesis and lipids synthesis in the cell. The C/N ratio is among the significant factors which helps microbes work optimally and it must be larger than 65 for maximal lipids storage capacity in *R. toruloides* [46]. This is primarily due to the suppression of glucose when the concentration of carbon sources is high in the media. This results in an osmotic shock as the Krebs cycle activity diminishes due to the intracellular rise in citric acid. The absorption of xylose gets reduced from lignocellulosic biomass hydrolysates. In this regard, cellodextrin transporters expressed in *S. cerevisiae* were investigated for their ability to alleviate this suppression caused by intracellular cellodextrin hydrolysis, allowing for improved utilization of the xylose contained in this form of hydrolysate [47].

A research was conducted on two concentrations of C/N (75.3 and 20.3) to measure the effect on the accumulation of biomass and lipids in *Rhodosporidium azoricum* and it was observed that yeast consumed the nitrogen source in 40 h, whereas with a low proportion, the compound remained present after 65 h of fermentation [48]. This variation was also noticed in a reduction in lipid synthesis (47%) when a minimal C/N ratio was applied, indicating that the strain's nitrogen restriction concerning the carbon supply had a detrimental effect on lipid formation. This is because nitrogen deficiency in the medium may have a detrimental impact on many essential enzymes, including 1,6-phosphate, 2-aldolase that

catalyzes the formation of dihydroxyacetone phosphate, a precursor to glyceraldehyde-3-P (required for TAGs synthesis). However, the absence of nitrogen in the medium promoted the action of glucose 6-phosphate-dehydrogenase [23]. This is advantageous because the pentose-phosphate pathway is the primary source of NADP, a critical intermediate for the proper activity of a ME involved in synthesizing lipids. Although few pathways have not been extensively studied in oleaginous species, some studies in *Yarrowia lipolytica* have revealed that the pentose-phosphate pathway is among the primary source of nicotinamide adenine dinucleotide phosphate oxidase (NADPH), which is responsible for this type of yeast's excessive lipid production [49].

This necessitates a more diligent approach to this parameter since the C/N ratio, carbohydrate (carbon source), and the fermentation technique (continuous, fed-batch, or batch) negatively influence lipids synthesis. The most remarkable results for *Rhodosporidium* strains were achieved when glucose was used as the only carbon source, and C/N rates in a range of 65 to 80, with lipids titer of up to 50% dry weight in batch systems and up to 61% dry weight in fed-batch systems [23]. In contrast to this, when mixes of carbohydrates like arabinose, xylose, and glucose are employed, the lipid yield decreases to nearly half when high C/N rates (65 and 80) are used; but, when a lower C/N (40) rate is employed, the lipid content can rise to 58% of dry weight.

The Stressful Conditions of Hydrolysate May Hinder Production of Lipids in *R. toruloides* as Observed During Osmotic Stress

One of the critical elements governing the development and synthesis of hydrolysates is the presence of chemicals that might cause yeast cells to experience osmotic stress. The above remark suggests a significant obstacle during the formulation from culture medium for competitive LPD synthesis. An excessive salt concentration in the broth causes harm to the yeast's cell wall owing to the high salt concentration. The latest research has demonstrated that osmotic stress significantly affects cell metabolism because it immediately regulates cell transcription stages such as mRNA polysomal association and translation inhibition. This parameter is crucial for choosing or enhancing the fermentation of broth constituents. In one of the studies, Singh et al. examined lipid synthesis in *R. toruloides* under high osmolarity induced by NaCl addition [50]. The findings indicated a direct relationship between glucose concentration and culture time. Similarly, the combination between glucose concentration and NaCl affects LPD synthesis. When the sodium chloride content in the medium was raised to 10% v/v, it had a detrimental effect on the formation of LPDs utilizing low glucose concentrations (5 g L1) [23]. This impact, however, is adaptable when glucose is added since the downside effect of sodium chloride is remunerated for by LPD formation. The phenomenon indicates that when the primary glucose concentration is greater than the NaCl concentration, it encourages a proportionate formation of LPDs concerning the NaCl added to the medium.

The Different Methods of Pretreatment of Lignocellulose Led to Variable pH Environment and Hence Will Impact Lipid Production

Rhodosporidium strains have shown a broad array of pH values (2.2–7.5), it was noticed that this factor has a major impact on biomass accumulation (10, 13.5, and 16 g L1, respectively), LPD titer (2.5, 4, and 6.4 g L1, respectively), and LPD productivity relative to biomass at low (pH = 4), intermediate (pH = 5.5), and high (pH = 7) values (32%, 38%,

and 27%, respectively) [51]. Yaegashi discovered that the pH of fermentation broth had a significant effect on the efficiency with which fermentable sugars were used for terpenes productivity from lignocellulosic hydrolysates, through a genetically engineered variant of *R. toruloides* [35]. They discovered that this strain produced the most terpene at low pH (3.4), suggesting that it could produce OAs in its original state. However, when the pH was reduced to 2.5, the strain's ability to use sugars was significantly impaired, suggesting that it was essential to maintain this pH range for effective sugar utilization (3.4 to 2.5). These values, however, may vary based on the substrate and the result. For example, the most incredible yields have been achieved with an average culture pH ranging between 4 and 7.5 [52, 53]. Nonetheless, this parameter affects biomass production and LPD yields, as small changes in pH values significantly affect the dissociation of certain toxic elements such as levulinic, formic, and acetic acids produced during the pretreatment and hydrolysis of hemicellulose and cellulose, resulting in cell membrane disruption and inhibition towards microbial growth.

The Nitrogen Source Scarcity and High Oxygen Conditions Produce More Lipids in *R. toruloides*

Naturally occurring oleaginous bacteria does not accumulate significant TAGs when the medium conditions are balanced. When a substrate is scarce (mostly nitrogen), bacteria use carbon supply to accumulate FAs as a defensive mechanism [23]. Limitation of nitrogen supply in the fermentation broth initiates a response that increases the AMP-deaminase enzyme, which in turn reduces the AMP levels inside microbe cells, resulting in a reduction of malate dehydrogenase enzyme seen in oleaginous yeast. Due to these considerations, a balance is required between the quantity of oxalacetate converted to citrate via aconitase and the amount of citrate malate translocase (CTM) mediated through the cytoplasm. Once citrate enters the cytosol, it is converted to AC and oxaloacetate by the ATP citrate lyase (ACL) (the first step in synthesizing FAs). On the other hand, recycling oxaloacetate is vital to maintaining a continual citrate supply. The enzymes malate dehydrogenase and citrate synthase are responsible for this [54]. However, achieving the LPD content using natural components with a minimal nitrogen concentration is challenging. Wu et al. used a variant of *R. toruloides* Y4 and reduced the phosphate and sulfur levels of the growth medium to generate LPDs, getting yields of 62% and 58%, respectively [55, 56].

Nonetheless, several investigations examined the influence of nitrogen sources in the microbial proliferation of the whole cell cycle. This impact has been examined in the context of nutrient restriction in the surroundings since detecting nutrients signals the exhaustion of nitrogen through the MAP and TOR kinase pathways. Relevant findings have been achieved in genetically modified strains that maintained their development rate in the presence of restricted nitrogen compared to natural variants. These findings imply that the growth process is adaptable and capable of separating cell division from the availability of nutrient elements in the surroundings [57]. This indicated that limiting these ingredients in fermentations for the generation of LPDs may be equally efficient and inexpensive in terms of nitrogen-rich materials.

This parameter is critical while fermenting carbohydrates into FAs. It has been proven that during fermentations at significant oxygen rates ranging from 100 to 225 rpm, LPD titers rise (from 9 to 46.3%) [23, 51]. This is because *Rhodosporidium* development is severely inhibited under low oxygen conditions since some pyruvate enters the fermentations through the enzyme pyruvate decarboxylase (PDC). The reduction in AC, formed

during the utilization of citrate through ATP citrate lyase, reflects those mentioned above. The scarcity of oxygen also results in a poor rate of cellular biomass formation [58]. All of this shows that oleaginous yeast requires a significant quantity of oxygen to create energy and perform optimal LPD biosynthesis. According to a previous study, keeping a dissolved oxygen concentration between 40 and 50% results in a high-density culture (107 g L1 dry biomass) and a significant LPD titer (67.5% w/w) [23]. Additionally, it is vital to keep tight control over the quantity of oxygen used throughout the fermentation process since its volume significantly affects the accumulation of unsaturated FAs. Another critical aspect in the effective utilization of carbohydrates is the presence of toxic compounds in the hydrolysate, which induce the carbon source to undergo partial or complete catabolism. This suppression reduces cell proliferation and LPD synthesis, and it can completely stop the microorganism.

The Fed Batch Process Led to High Lipid Production Compared to Other Cultivation Processes

Different microbial strains are used for cultivation in lignocellulosic hydrolysates since their lipid concentration varies from 2 to 39.5 g/L and dry cell weight varies from 17.27 to 61.3% [59]. Fed-batch mode is more efficient in terms of productivity than other modes. *R. toruloides* AS 2.1389 strain has shown a lipid productivity rate of 0.16 (g/L/h) and 0.025 (g/L/h) when employed under batch mode by Brandenburg et al. (2021) [60] and Carmona-Cabello et al. (2021) [61]. Lakshmidevi et al. (2021) [62] studied CCT 0783 strain for the saccharified sweet sorghum juice (20%) feedstock with a yield of 0.105 (g LPD/g carbon) and productivity of 0.057 (g/L/h) in batch fermentation. Monteiro de Oliveira et al. [63] and Zheng et al. (2021) [5] used a fed-batch strategy for the fermentation of *R. toruloides* DSM 4444 and it showed a yield of 0.17 (g lipid/g carbon) and efficacy of 0.17 (g/L/h) when crude glycerol of concentration 50 g/L was used a carbon source [63]. *R. toruloides* Y4 strain is observed to produce 0.28 (g lipid/g carbon) of LPD utilizing glucose as the carbon source [5].

Genetic Manipulation Approaches for Enhancement in Fatty Acid Production in *R. toruloides*

Zhao and colleagues sequenced the genome of *R. toruloides NP11* and discovered a 20.2 Mb genome with a GC (Guanine-Cytocine) content of 61.9% and 8171 protein-coding genes [28]. Since then, numerous *Rhodosporidium species* draft genomes have been reported [64]. Functional and comparative genotyping can disclose the genetic architecture, protein-coding domains, speculated peptides, and functioning RNA of different species.

The combination of transcriptomics and genomics enabled the identification of introns of genomic composition and fatty acid synthase system involved in metabolic pathways [28]. In other investigations, transcriptome analysis was utilized to characterize the gene expression patterns and global responses of *R. toruloides* to a variety of stress conditions or the usage of a diverse range of carbon sources [24, 65–67]. These efforts resulted in the identification of many stress-resistant genes and regulatory chromosomes. However, proteomic analysis was used while assessing the peculiarities in protein expression across a range of circumstances. This approach is critical because it provides information to identify proteins implicated in lipid synthesis [3]. Additionally, it aids in investigating xylose

metabolism [68] and the identification of dynamic changes in essential proteins linked with lipids droplets [69]. However, unlike standalone proteomic or transcriptomic analysis, combined proteomic and transcriptomic research provides additional evidence for identifying critical genes and pathways [67].

Different genetic approaches have enhanced the generation titers, efficiencies, and economic productivity of fatty acids and carotenoids in *R. toruloides*. In *R. toruloides*, molecular genetic efforts were made to enhance the quality of fatty acids synthesis with complete genome sequencing and the development of genomic technologies, as indicated in Table 2. Zhang and colleagues over-expressed acetyl-CoA carboxylase (ACC1) and diacylglycerol acyltransferase genes for enhanced fatty acids accumulation by 1.95 × (16 g/L of lipid content from sugars in shake flask culture) in *R. toruloides* [70]. Employing glucose in fed-batch culture, the same research team changed the microorganism by inducing the release of stearoyl-CoA desaturase (SCD) and discovered a coefficient of 1.42 enhancements in lipid concentration.

Challenges for Downstream Processing of Triglycerides Accumulated by *R. toruloides*

Extraction of Lipids

Although oleaginous microbes seem attractive for the fact that 20 to 80% of dry weight is lipids, but there are a number of challenges leading to high cost of production. One of the major challenges during extraction of lipids arises due to large quantity of liquid broth. The problem is due to complicated protocols which are mostly focused on dry biomass. The different solvents are used for passing through the cell membrane of yeast, but the presence of liquid culture hinders solubility problems. Furthermore, traditional method of drying microbial biomass using freeze drying or oven drying makes energy intensive and costly. Removal of moisture using harvesting and dewatering techniques like centrifugation require lot of energy but can reduce moisture content up to 90% [75].

One of the research groups in 2022 developed a novel strategy for extraction of lipids from *R. toruloids* biomass. The protocol include use of either hexane or sunflower oils solvents leading to elimination of harvesting techniques to disrupt yeast cells. The cells were disrupted using high pressure homogenization (HPH) and 10% sunflower oil or hexane was added followed by centrifugation. After overnight mixing at room temperature, the hexane or oil mixture was separated using separating funnel. This method was easy, inexpensive, and environmentally friendly. The protocol found to be successful as resulted in higher outcomes achieving 55% of total lipids [76].

Transesterification

Following extraction, the triglycerides must undergo a procedure called transesterification, which will decrease their fluidity and turn them into long chains of alkyl esters often categorized as biodiesel [74]. The fatty acid combines with alcohol, like ethanol and methanol, in the vicinity of a catalyst during the transesterification process to generate fatty acid methyl esters and glycerol. The process may be classed as a chemically catalyzed reaction, a non-catalyzed reaction, or an enzyme-catalyzed reaction according to the method of catalysis. Heterogeneous and homogeneous reactions are subcategories of chemically

Table 2 Genetic engineered *R. toruloides* and its respective outcomes

S. no	Strain	Gene manipulated/inserted	Yield/production	Ref
1	*R. toruloides* IFO0880	Acetyl-CoA carboxylase (*ACC1)* and Diacylglycerol acyltransferase (*DGA1)*	Metabolic engineered strain of R. toruloides for the overexpression of *ACCI* and *DGA1* genes resulted in 16.4 g/L and 9.5 g/L of FA from 70 g/L glucose and xylose respectively	[71, 72]
2	*R. toruloides*	Elongase-1 (*ELO1*) and Delta-12 fatty acid desaturase (*FAD2*)	Deletion of *FAD2* and overexpression of *ELO1* genes resulted in increase in oleic content up to 23%	[71]
3	*R. toruloides* NP11	Delta-12 fatty acid desaturase (*FAD2*)	Overexpression of *FAD2* increased the accumulation of linolenic acid by 5 times and quantitatively reached to 1.3 g/L	[8]
4	*R. toruloides* CECT 13085	3-ketoacyl-CoA synthases (*KCS*)	Heterologous expression of *KCS* compelled the strain to produce very long fatty acids (VLFAs) such as erucic acid (5.8 g/L) and nervonic acid (7.9 g/L)	[73]
5	*R. toruloides*	Aldehyde dehydrogenase (*ALD1*)	Inactivation of *ALD1* lead to 1.2 times production of Linolenic acid and 49% increase in fatty acid accumulation	[8]
6	*R. toruloides* CECT 13085	Fatty acyl-CoA reductase (*FAC*)	Overexpression of *FAC* resulted in higher production of C16-C18 fatty alcohols, quantitatively up to 8 g/L	[71]
7	*R. toruloides* IFO0880	Malic enzyme (ME)	Increased expression of *ME* genes resulted in higher LPD production to 89.4 g/L	[8, 71]
8	*R. toruloides* IFO0880	Stearoyl-CoA desaturase (*SCD*)	Reported to increase the LPD titer by a factor of 1.2 from glucose carbon source in fed-batch growth	[72]
9	*R. toruloides* IFO0880	Bisabolene synthase (*BIS*)	Optimized synthesis and insertion of *BIS* genes resulted in production of 680 mg/L of Bisabolene	[71]
10	*R. toruloides* NP11	Phosphotransacetylase (*PTA*)	Metabolic constructed strain was found to enhance the cell mass by 8.5% (12.8 g/L) and LPD content up to 64.4% (1.20 g/L/d) in comparison to parent strain	[74]

mediated reactions. Acidic or basic mediators in solid forms, such as calcium-based metal oxides and sulfated zirconia, are used to facilitate the heterogeneous response [77]. Parallel to this, the homogeneous effect proceeds when an acidic or alkali catalyst is present in a form of liquid. The instances of such catalysts are HCl, H2SO4, NaOH, and KOH [78]. While the alkali-based transesterification approach proceeds swiftly, it is not practical for biomasses with greater free fatty acid levels because of soap formation during this process. In contrast to this, the drawbacks of acid-based transesterification include a prolonged incubation period, corrosive character, greater power usage, increased glycerin refining expenses, etc.

During supercritical circumstances, the non-catalyzed transesterification process occurs swiftly with the product may be readily extracted without producing waste. The method's drawbacks include the pricey maximum temp ranging from 250 to 400 °C and pressure between 10 and 30 MPa requirements [79]. The study examined enzyme transesterification and highlighted that the reaction occurring under benign circumstances has good selectivity. This enzymatic transesterification solves the issues with chemical-catalyzed and non-catalyzed processes [80]. Shimada et al. (2002) [81] showed that immobilized lipase from Candida Antarctica converted docosahexaenoic acid into fatty acid methyl eaters at a rate of 98.5%.

Conversion of Fatty Acids to Biofuels

Many commercial operations have been discovered in the oleaginous yeast. Lipids, one of the crucial metabolites from oleaginous yeast, were undertaken for the formation of drop-in-alkanes and biodiesel. Although many substitute forms of renewable energy can solve upcoming power generation requirements, organic matter promises an easily achieved, low-cost remedy to a transportation fuel for replacing traditional diesel [82]. This can be achieved by thermal treatment through sugars pyrolysis/ hydrodeoxygenation (HDO) or lipids transesterification to drop-in-alkanes and biodiesel, respectively (Fig. 3) [83]. Hydrotreating or transesterification of oleaginous TAGs results in biofuels with substantially

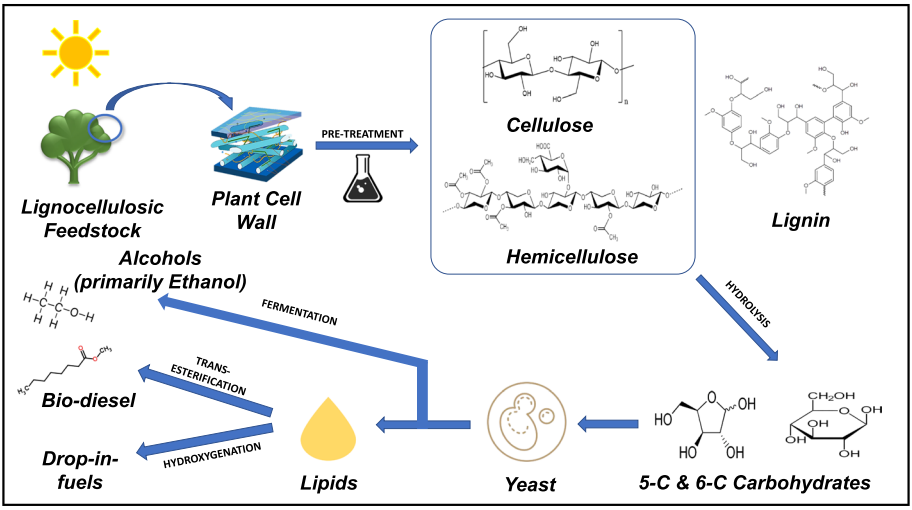


Fig. 3 Schematic representation of biofuels production from lignocellulosic biomass using *R. toruloides*

various characteristics. Hydrotreatment of TAGs can follow one of two routes: hydrodeoxygenation (HDO) to accumulate n-alkanes, propane, and water; or hydro-decarboxylation (HDC) to yield one less carbon atom hydrocarbon from the initial triglycerides, resulting in carbon monoxide, carbon dioxide, and propane as by-products [82]. Hydrotreating may break the fatty acid chains to obtain high-grade biofuels. Although this is beneficial for existing petrochemical industries, it demands energy-intensive working temperatures of 250–350 °C and a sustainable hydrogen source which is a major hurdle for localized fuel generation in developing countries. On the other hand, commercial biodiesel is now made by transesterification of C14–C20 TAGs components of fatty acids with C1–C2 alcohols [84, 85] to produce FAMEs, which are then converted into biodiesel. Hence, biodiesel is a potential alternative to conventional petroleum-derived diesel [86]. Nevertheless, fuel purity is strongly reliant on the oil supply, with longer chain length (> 18 carbon) molecules lowering fuel qualities like cloud and pour points. Drop-in oil has high energy density and they have the potential to be used directly as an alternative to diesel, gasoline, and jet fuel such as alkanes. They have advantages over petroleum, i.e., contain no sulfur and polyaromatic hydrocarbon; it does not generate particulate matter and carbon monoxides when burning [87].

Conclusions

The presence of sustainable biomasses like lignocellulosic materials and an enhancement in the utilization of carbon sources generated from this raw material has intensified the investigation into the aggressive metabolization of feedstock from *Rhodosporidium* isolates. Our study consists of several aspects: (i) the metabolism involved for fatty acids accumulation; (ii) optimization of culture parameters and nutrient sources for better growth and derivatives; and (iii) the advancements in genetic engineering to permit the maximum yield of lipid. Regarding *R. toruloides*' applicability, there could be more advancements in the synthesis of carotenoids and triglycerides, mainly via metabolomics, which will be wholly harnessed in the pathways leading to such derivatives. New industrial applicable products will be determined in the coming years, as with the synthesis of new and more practical components not natively formed by *R. toruloides*, such as FAMEs or fatty alcohols. Several oleic-derived chemicals, notably those formed from both the routes (lipids and carotenoid biosynthesis pathways) with specific physiological stasis in the yeast. The *R. toruloides* bioprocesses might be made more commercially viable by yielding several high-value additives. It may be envisioned as an industrialized yeast with significant potential in the production of goods and biofuels in the coming years, with the aid of bioengineering and biotechnological engineering, which might set the stage for a worldwide green economy.

Author Contribution RK did the conceptualization for this manuscript. Writing and review literature was done by SS, AG. The final editing was done by RK and RR. All authors read and approved the manuscript.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest The authors declare no competing interests.

References

1. d'Espaux, L., Mendez-Perez, D., Li, R., & Keasling, J. D. (2015). Synthetic biology for microbial production of lipid-based biofuels. *Current Opinion in Chemical Biology, 29*, 58–65.
2. Zhou, Y. J., Kerkhoven, E. J., & Nielsen, J. (2018). Barriers and opportunities in bio-based production of hydrocarbons. *Nature Energy, 3*(11), 925–935.
3. Wen, Z., Zhang, S., Odoh, C. K., Jin, M., & Zhao, Z. K. (2020). Rhodosporidium toruloides - A potential red yeast chassis for lipids and beyond. *FEMS Yeast Research, 20*(5), 38.
4. Osorio-González, C. S., Hegde, K., Brar, S. K., Kermanshahipour, A., & Avalos-Ramírez, A. (2019). Challenges in lipid production from lignocellulosic biomass using Rhodosporidium sp.; A look at the role of lignocellulosic inhibitors. *Biofuels, Bioproducts and Biorefining, 13*(3), 740–759.
5. Zheng, X., Hu, R., Chen, D., Chen, J., He, W., Huang, L., … Xue, T. (2021). Lipid and carotenoid production by the Rhodosporidium toruloides mutant in cane molasses. *Bioresource Technology*, *326*, 124816.
6. Zhao, X., Peng, F., Du, W., Liu, C., & Liu, D. (2012). Effects of some inhibitors on the growth and lipid accumulation of oleaginous yeast Rhodosporidium toruloides and preparation of biodiesel by enzymatic transesterification of the lipid. *Bioprocess and Biosystems Engineering, 35*(6), 993–1004.
7. SànchezNogué, V., Black, B. A., Kruger, J. S., Singer, C. A., Ramirez, K. J., Reed, M. L., … Beckham, G. T. (2018). Integrated diesel production from lignocellulosic sugars via oleaginous yeast. *Green Chemistry*, *20*(18), 4349–4365.
8. Park, Y. K., Nicaud, J. M., & Ledesma-Amaro, R. (2018). The engineering potential of Rhodosporidium toruloides as a workhorse for biotechnological applications. *Trends in Biotechnology, 36*(3), 304–317.
9. Yu, Y., & Shi, S. (2023). Development and perspective of Rhodotorula toruloides as an efficient cell factory. *Journal of Agricultural and Food Chemistry, 71*(4), 1802–1819.
10. Singh, G., Sinha, S., Bandyopadhyay, K. K., Lawrence, M., & Paul, D. (2018). Triauxic growth of an oleaginous red yeast Rhodosporidium toruloides on waste "extract" for enhanced and concomitant lipid and β-carotene production. *Microbial Cell Factories, 17*(1), 1–10.
11. Dai, X., Shen, H., Li, Q., Rasool, K., Wang, Q., Yu, X., … Zhao, Z. K. (2019). Microbial lipid production from corn stover by the oleaginous yeast Rhodosporidium toruloides using the PreSSLP process. *Energies 12*(6), 1053.
12. Keskin, A., Ünlü, A. E., & Takaç, S. (2023). Utilization of olive mill wastewater for selective production of lipids and carotenoids by Rhodotorula glutinis. *Applied Microbiology and Biotechnology, 107*, 4973–4985.
13. Into, P., Pontes, A., Sampaio, J. P., & Limtong, S. (2020). Yeast diversity associated with the phylloplane of corn plants cultivated in Thailand. *Microorganisms, 8*(1), 80.
14. Buck, J. W., & Andrews, J. H. (1999). Attachment of the yeast Rhodosporidium toruloides is mediated by adhesives localized at sites of bud cell development. *Applied and Environmental Microbiology, 65*(2), 465–471.
15. Gadanho, M., & Sampaio, J. P. (2005). Occurrence and diversity of yeasts in the mid-atlantic ridge hydrothermal fields near the Azores Archipelago. *Microbial Ecology, 50*(3), 408–417.
16. Gadanho, M., Libkind, D., & Sampaio, J. P. (2006). Yeast diversity in the extreme acidic environments of the Iberian Pyrite Belt. *Microbial Ecology, 52*(3), 552–563.
17. Wang, Q. M., Yurkov, A. M., Göker, M., Lumbsch, H. T., Leavitt, S. D., Groenewald, M., … Bai, F. Y. (2015). Phylogenetic classification of yeasts and related taxa within Pucciniomycotina. *Studies in Mycology*, *81*, 149–189.
18. Bonturi, N., Crucello, A., Viana, A. J. C., & Miranda, E. A. (2017). Microbial oil production in sugarcane bagasse hemicellulosic hydrolysate without nutrient supplementation by a Rhodosporidium toruloides adapted strain. *Process Biochemistry, 57*, 16–25.
19. Jagtap, S. S., & Rao, C. V. (2018). Production of d-arabitol from d-xylose by the oleaginous yeast Rhodosporidium toruloides IFO0880. *Applied Microbiology and Biotechnology, 102*(1), 143–151.
20. Yang, X., Jin, G., Wang, Y., Shen, H., & Zhao, Z. K. (2015). Lipid production on free fatty acids by oleaginous yeasts under non-growth conditions. *Bioresource Technology, 193*, 557–562.
21. Bertacchi, S., Bettiga, M., Porro, D., & Branduardi, P. (2020). Camelina sativa meal hydrolysate as sustainable biomass for the production of carotenoids by Rhodosporidium toruloides. *Biotechnology for Biofuels, 13*(1), 1–10.
22. Lopes, H. J. S., Bonturi, N., & Miranda, E. A. (2020). Rhodotorula toruloides single cell oil production using eucalyptus urograndis hemicellulose hydrolysate as a carbon source. *Energies, 13*(4), 795.

23. Osorio-González, C. S., Hegde, K., Ferreira, P., Brar, S. K., Kermanshahipour, A., Soccol, C. R., & Avalos-Ramírez, A. (2019). Lipid production in Rhodosporidium toruloides using C-6 and C-5 wood hydrolysate: A comparative study. *Biomass and Bioenergy*, *130*.
24. Wang, Y., Zhang, S., Zhu, Z., Shen, H., Lin, X., Jin, X., … Zhao, Z. K. (2018). Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast Rhodosporidium toruloides. *Biotechnology for Biofuels*, *11*(1), 1–15.
25. Czabany, T., Wagner, A., Zweytick, D., Lohner, K., Leitner, E., Ingolic, E., & Daum, G. (2008). Structural and biochemical properties of lipid particles from the yeast Saccharomyces cerevisiae*. *Journal of Biological Chemistry, 283*(25), 17065–17074.
26. Athenstaedt, K., Jolivet, P., Boulard, C., Zivy, M., Negroni, L., Nicaud, J. M., & Chardot, T. (2006). Lipid particle composition of the yeast Yarrowia lipolytica depends on the carbon source. *Proteomics, 6*(5), 1450–1459.
27. Koivuranta, K., Castillo, S., Jouhten, P., Ruohonen, L., Penttilä, M., & Wiebe, M. G. (2018). Enhanced triacylglycerol production with genetically modified Trichosporon oleaginosus. *Frontiers in Microbiology, 9*(JUN), 1337.
28. Zhu, Z., Zhang, S., Liu, H., Shen, H., Lin, X., Yang, F., … Zhao, Z. K. (2012). A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides. *Nature Communications*, *3*(1), 1–12.
29. Chattopadhyay, A., & Maiti, M. K. (2021). Lipid production by oleaginous yeasts. *Advances in Applied Microbiology, 116*, 1–98.
30. Zhang, Y., Adams, I. P., & Ratledge, C. (2007). Malic enzyme: The controlling activity for lipid production? Overexpression of malic enzyme in Mucor circinelloides leads to a 2.5-fold increase in lipid accumulation. *Microbiology, 153*(7), 2013–2025.
31. Huang, Y. B., & Fu, Y. (2013). Hydrolysis of cellulose to glucose by solid acid catalysts. *Green Chemistry, 15*(5), 1095–1111.
32. Mahalingam, L., Abdulla, R., Sani, S. A., Sabullah, M. K., Faik, A. A. M., & Misson, M. (2022). Lignocellulosic biomass – A sustainable feedstock for acetone-butanol-ethanol fermentation. *Periodica Polytechnica Chemical Engineering, 66*(2), 279–296.
33. Slininger, P. J., Dien, B. S., Kurtzman, C. P., Moser, B. R., Bakota, E. L., Thompson, S. R., … Dale, B. E. (2016). Comparative lipid production by oleaginous yeasts in hydrolyzates of lignocellulosic biomass and process strategy for high titers. *Biotechnology and Bioengineering*, *113*(8), 1676–1690.
34. Matsakas, L., Bonturi, N., Miranda, E. A., Rova, U., & Christakopoulos, P. (2015). High concentrations of dried sorghum stalks as a biomass feedstock for single cell oil production by Rhodosporidium toruloides. *Biotechnology for Biofuels*, *8*(1).
35. Yaegashi, J., Kirby, J., Ito, M., Sun, J., Dutta, T., Mirsiaghi, M., … Gladden, J. M. (2017). Rhodosporidium toruloides: A new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotechnology for Biofuels*, *10*(1), 1–13.
36. Gong, Z., Wang, X., Yuan, W., Wang, Y., Zhou, W., Wang, G., & Liu, Y. (2020). Fed-batch enzymatic hydrolysis of alkaline organosolv-pretreated corn stover facilitating high concentrations and yields of fermentable sugars for microbial lipid production. *Biotechnology for Biofuels, 13*(1), 1–15.
37. Jędrzejczyk, M., Soszka, E., Czapnik, M., Ruppert, A. M., & Grams, J. (2019). Physical and chemical pretreatment of lignocellulosic biomass. *Second and Third Generation of Feedstocks: The Evolution of Biofuels*, 143–196.
38. Jönsson, L. J., & Martín, C. (2016). Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresource Technology, 199*, 103–112.
39. Nanda, S., Mohammad, J., Reddy, S. N., Kozinski, J. A., & Dalai, A. K. (2014). Pathways of lignocellulosic biomass conversion to renewable fuels. *Biomass Conversion and Biorefinery, 4*(2), 157–191.
40. Zhang, H., Han, L., & Dong, H. (2021). An insight to pretreatment, enzyme adsorption and enzymatic hydrolysis of lignocellulosic biomass: Experimental and modeling studies. *Renewable and Sustainable Energy Reviews, 140*, 110758.
41. Świątek, K., Gaag, S., Klier, A., Kruse, A., Sauer, J., & Steinbach, D. (2020). Acid hydrolysis of lignocellulosic biomass: Sugars and furfurals formation. *Catalysts, 10*(4), 437.
42. Althuri, A., Chintagunta, A. D., Sherpa, K. C., & Banerjee, R. (2018). Simultaneous saccharification and fermentation of lignocellulosic biomass, 265–285.
43. Adrio, J. L. (2017). Oleaginous yeasts: Promising platforms for the production of oleochemicals and biofuels. *Biotechnology and Bioengineering, 114*(9), 1915–1920.
44. Wiebe, M. G., Koivuranta, K., Penttilä, M., & Ruohonen, L. (2012). Lipid production in batch and fed-batch cultures of Rhodosporidium toruloides from 5 and 6 carbon carbohydrates. *BMC Biotechnology*, *12*, 26.

45. Chandel, A. K., da Silva, S. S., & Singh, O. V. (2013). Detoxification of lignocellulose hydrolysates: Biochemical and metabolic engineering toward white biotechnology. *Bioenergy Research, 6*(1), 388–401.
46. Jin, M., Slininger, P. J., Dien, B. S., Waghmode, S., Moser, B. R., Orjuela, A., Sousa, L. C., & Balan, V. (2015). Microbial lipid-based lignocellulosic biorefinery: Feasibility and challenges. *Trends in Biotechnology, 33*(1), 43–54.
47. Jin, Y. S., & Cate, J. H. (2017). Metabolic engineering of yeast for lignocellulosic biofuel production. *Current Opinion in Chemical Biology, 41*, 99–106.
48. Capusoni, C., Rodighiero, V., Cucchetti, D., Galafassi, S., Bianchi, D., Franzosi, G., & Compagno, C. (2017). Characterization of lipid accumulation and lipidome analysis in the oleaginous yeasts Rhodosporidium azoricum and Trichosporon oleaginosus. *Bioresource Technology, 238*, 281–289.
49. Wasylenko, T. M., Ahn, W. S., & Stephanopoulos, G. (2015). The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. *Metabolic Engineering, 30*, 27–39.
50. Singh, G., Jawed, A., Paul, D., Bandyopadhyay, K. K., Kumari, A., & Haque, S. (2016). Concomitant production of lipids and carotenoids in Rhodosporidium toruloides under osmotic stress using response surface methodology. *Frontiers in Microbiology, 7*(OCT), 1686.
51. Jiru, T. M., Groenewald, M., Pohl, C., Steyn, L., Kiggundu, N., & Abate, D. (2017). Optimization of cultivation conditions for biotechnological production of lipid by Rhodotorula kratochvilovae (syn, Rhodosporidium kratochvilovae) SY89 for biodiesel preparation. *3 Biotech, 7*(2), 1–11.
52. Polburee, P., Yongmanitchai, W., Lertwattanasakul, N., Ohashi, T., Fujiyama, K., & Limtong, S. (2015). Characterization of oleaginous yeasts accumulating high levels of lipid when cultivated in glycerol and their potential for lipid production from biodiesel-derived crude glycerol. *Fungal Biology, 119*(12), 1194–1204.
53. Uprety, B. K., Dalli, S. S., & Rakshit, S. K. (2017). Bioconversion of crude glycerol to microbial lipid using a robust oleaginous yeast Rhodosporidium toruloides ATCC 10788 capable of growing in the presence of impurities. *Energy Conversion and Management, 135*, 117–128.
54. Kosa, M., & Ragauskas, A. J. (2011). Lipids from heterotrophic microbes: Advances in metabolism research. *Trends in Biotechnology, 29*(2), 53–61.
55. Wu, S., Hu, C., Jin, G., Zhao, X., & Zhao, Z. K. (2010). Phosphate-limitation mediated lipid production by Rhodosporidium toruloides. *Bioresource Technology, 101*(15), 6124–6129.
56. Wu, S., Zhao, X., Shen, H., Wang, Q., & Zhao, Z. K. (2011). Microbial lipid production by Rhodosporidium toruloides under sulfate-limited conditions. *Bioresource Technology, 102*(2), 1803–1807.
57. Navarro, F. J., Weston, L., & Nurse, P. (2012). Global control of cell growth in fission yeast and its coordination with the cell cycle. *Current Opinion in Cell Biology, 24*(6), 833–837.
58. Zhao, X., Wu, S., Hu, C., Wang, Q., Hua, Y., & Zhao, Z. K. (2010). Lipid production from Jerusalem artichoke by Rhodosporidium toruloides Y4. *Journal of Industrial Microbiology and Biotechnology, 37*(6), 581–585.
59. Zhao, Y., Song, B., Li, J., & Zhang, J. (2022). Rhodotorula toruloides: An ideal microbial cell factory to produce oleochemicals, carotenoids, and other products. *World Journal of Microbiology and Biotechnology*. Springer.
60. Brandenburg, J., Blomqvist, J., Shapaval, V., Kohler, A., Sampels, S., Sandgren, M., & Passoth, V. (2021). Oleaginous yeasts respond differently to carbon sources present in lignocellulose hydrolysate. *Biotechnology for Biofuels, 14*(1), 1–12.
61. Carmona-Cabello, M., García, I. L., Papadaki, A., Tsouko, E., Koutinas, A., Dorado, M. P. (2021). Biodiesel production using microbial lipids derived from food waste discarded by catering services. *Bioresource Technology, 323*, 124597.
62. Lakshmidevi, R., Ramakrishnan, B., Ratha, S. K., Bhaskar, S., & Chinnasamy, S. (2021). Valorisation of molasses by oleaginous yeasts for single cell oil (SCO) and carotenoids production. *Environmental Technology and Innovation, 21*, 101281.
63. Monteiro de Oliveira, P., Aborneva, D., Bonturi, N., & Lahtvee, P. J. (2021). Screening and growth characterization of non-conventional yeasts in a hemicellulosic hydrolysate. *Frontiers in Bioengineering and Biotechnology, 9*, 313.
64. Tran, T. N., Ngo, D. H., Nguyen, N. T., & Ngo, D. N. (2019). Draft genome sequence data of Rhodosporidium toruloides VN1, a strain capable of producing natural astaxanthin. *Data in Brief*, *26*, 104443.
65. Coradetti, S. T., Pinel, D., Geiselman, G. M., Ito, M., Mondo, S. J., Reilly, M. C., … Skerker, J. M. (2018). Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. *eLife*, *7*.

66. Lu, L., Wang, J., Zhu, R., Lu, H., Zheng, X., Yu, T. (2015). Transcript profiling analysis of Rhodosporidium paludigenum-mediated signalling pathways and defense responses in mandarin orange. *Food Chemistry, 172*, 603–612.
67. Qi, F., Zhao, X., Kitahara, Y., Li, T., Ou, X., Du, W., … Huang, J. (2017). Integrative transcriptomic and proteomic analysis of the mutant lignocellulosic hydrolyzate-tolerant Rhodosporidium toruloides. *Engineering in Life Sciences*, *17*(3), 249–261.
68. Tiukova, I. A., Brandenburg, J., Blomqvist, J., Sampels, S., Mikkelsen, N., Skaugen, M., … Kerkhoven, E. J. (2019). Proteome analysis of xylose metabolism in Rhodotorula toruloides during lipid production. *Biotechnology for Biofuels*, *12*(1), 1–17.
69. Zhu, Z., Ding, Y., Gong, Z., Yang, L., Zhang, S., Zhang, C., … Zhaoa, Z. K. (2015). Dynamics of the lipid droplet proteome of the oleaginous yeast Rhodosporidium toruloides. *Eukaryotic Cell*, *14*(3), 252–264.
70. Zhang, S., Skerker, J. M., Rutter, C. D., Maurer, M. J., Arkin, A. P., & Rao, C. V. (2016). Engineering Rhodosporidium toruloides for increased lipid production. *Biotechnology and Bioengineering, 113*(5), 1056–1066.
71. Saini, R., Hegde, K., Brar, S. K., & Vezina, P. (2020). Advanced biofuel production and road to commercialization: An insight into bioconversion potential of Rhodosporidium sp. *Biomass and Bioenergy, 132*, 105439.
72. Zhang, S., Ito, M., Skerker, J. M., Arkin, A. P., & Rao, C. V. (2016). Metabolic engineering of the oleaginous yeast Rhodosporidium toruloides IFO0880 for lipid overproduction during high-density fermentation. *Applied Microbiology and Biotechnology, 100*(21), 9393–9405.
73. Fillet, S., Ronchel, C., Callejo, C., Fajardo, M. J., Moralejo, H., & Adrio, J. L. (2017). Engineering Rhodosporidium toruloides for the production of very long-chain monounsaturated fatty acid-rich oils. *Applied Microbiology and Biotechnology, 101*(19), 7271–7280.
74. Sampath Kumar, N. S., Sruthi Geetha, R., Thirumala Vasu, A., & Vijaya Ramu, D. (2019). Performance assessment of biodiesel transesterified from labeo rohita visceral oil using IC engine. *Current Trends in Biotechnology and Pharmacy, 13*(4), 348–356.
75. Samarasinghe, N. (2012). Effect of high pressure homogenization on aqueous phase solvent extraction of lipids from nannochloris oculata microalgae. *Journal of Energy and Natural Resources, 1*, 1.
76. Dias, C., Nobre, B. P., Santos, J. A. L., Lopes Da Silva, T., & Reis, A. (2022). Direct lipid and carotenoid extraction from Rhodosporidium toruloides broth culture after high pressure homogenization cell disruption: Strategies, methodologies, and yields. *Biochemical Engineering Journal, 189*, 108712.
77. Aransiola, E. F., Ojumu, T. V., Oyekola, O. O., Madzimbamuto, T. F., & Ikhu-Omoregbe, D. I. O. (2014). A review of current technology for biodiesel production: State of the art. *Biomass and Bioenergy, 61*(276), 297.
78. Ong, H. R., Khan, M. R., Chowdhury, M. N. K., Yousuf, A., & Cheng, C. K. (2014). Synthesis and characterization of CuO/C catalyst for the esterification of free fatty acid in rubber seed oil. *Fuel, 120*, 195–201.
79. Stamenković, O. S., Veličković, A. V., & Veljković, V. B. (2011). The production of biodiesel from vegetable oils by ethanolysis: Current state and perspectives. *Fuel, 90*(11), 3141–3155.
80. Christopher, L. P., Kumar, H., & Zambare, V. P. (2014). Enzymatic biodiesel: Challenges and opportunities. *Applied Energy, 119*, 497–520.
81. Shimada, Y., Watanabe, Y., Sugihara, A., & Tominaga, Y. (2002). Enzymatic alcoholysis for biodiesel fuel production and application of the reaction to oil processing. *Journal of Molecular Catalysis - B Enzymatic, 17*(3–5), 133–142.
82. Lee, A. F., & Wilson, K. (2015). Recent developments in heterogeneous catalysis for the sustainable production of biodiesel. *Catalysis Today, 242*(Part A), 3–18.
83. Lin, C. S. K., Pfaltzgraff, L. A., Herrero-Davila, L., Mubofu, E. B., Abderrahim, S., Clark, J. H., … Luque, R. (2013). Food waste as a valuable resource for the production of chemicals, materials and fuels. Current situation and global perspective. *Energy & Environmental Science*, *6*(2), 426–464.
84. Donnis, B., Egeberg, R. G., Blom, P., & Knudsen, K. G. (2009). Hydroprocessing of bio-oils and oxygenates to hydrocarbons. Understanding the reaction routes. *Topics in Catalysis, 52*(3), 229–240.
85. Othman, M. R., Helwani, Z., Martunus, & Fernando, W. J. N. (2009). Synthetic hydrotalcites from different routes and their application as catalysts and gas adsorbents: A review. *Applied Organometallic Chemistry, 23*(9), 335–346.
86. Knothe, G. (2010). Biodiesel: Current trends and properties. *Topics in Catalysis, 53*(11–12), 714–720.

87. Xue, S. J., Chi, Z., Zhang, Y., Li, Y. F., Liu, G. L., Jiang, H., … Chi, Z. M. (2018). Fatty acids from oleaginous yeasts and yeast-like fungi and their potential applications. *Critical Reviews in Biotechnology*, *38*(7), 1049–1060.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Power Quality Enhancement using Adaptive Notch Filter in Grid Distribution System

Chirag Singh
Department of Electrical Engineering
Delhi Technological University (DTU)
imchirag.2717@gmail.com

Gautam Kumar
Department of Electrical Engineering
Delhi Technological University (DTU)
gautamkumar.dtu@gmail.com

Himanshu Kumar
Department of Electrical Engineering
Delhi Technological University (DTU)
himanshu.k2899@gmail.com

Ankita Arora
Department of Electrical Engineering
Delhi Technological University (DTU)
aroraankita985@gmail.com

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205688

***Abstract*— In this work three different techniques for harmonic current compensation and extraction utilizing a VSC (voltage source converter) in a 3-phase, three-wire distribution system are described. The stationary load current components were calculated by Synchronous Reference Frame Theory (SRFT), which has its foundation on the transformations of load currents proposed by Clarke and Park. A Notch Filter is a band reject (band stop) filter that allows the majority of frequencies to pass through unaltered while attenuating select frequencies to very low values in a particular range. The basic portion of the contaminated load current is fetched using a notch filter, and VSC provides the appropriate correction. The filter that is self-tuned and also separates the active part of load current is called an adaptive notch filter. These suggested methods are straightforward and can quickly identify and reduce harmonic current using shunt compensation. We have used MATLAB Simulink to evaluate, analyze, compare and simulate the techniques.**

Keywords*— Notch Filter; Adaptive Notch; SRFT; Fundamental Current; THD; DSTATCOM.

I. INTRODUCTION

Power quality, also known as 'power quality health' is the process of estimation by which we get to know how steady and consistent the electrical system is. Only by diagnosing and solving power quality issues can help a power distribution organization save costs by improving energy use and preventing further equipment damage. Power quality (PQ) issues affecting electrical power distribution systems include inadequate power factor, unbalanced loads, harmonics, voltage distortion, etc. In power distribution systems, nonlinear loads with power electronics components, such as lighting loads (LED, CFL), VFD that is Variable Frequency Drives and SMPS which are Switch Mode Power Supplies make a significant contribution. Higher PQ issues such as the introduction of harmonics into the voltage and current etc. result from this. It is crucial to identify solutions to address power quality issues since their continuing existence degrades the standard of power at the consumer level. Primary effects of poor quality of power distribution consists swells and dips in the voltage levels, harmonic introduction, voltage and current unbalancing, flickering, etc. Conventionally, capacitors, reactors, capacitor banks, tap changing transformers are installed to tackle and improve PQ issues in distribution systems. Due to slow compensation and lack of active load compensation of these methods, new power devices are being designed and developed actively. One such technology, the Shunt Active Power Filter (SAPF), which combines a DC Link Capacitor and a Voltage Source Converter (VSC), seeks to reduce a number of PQ issues. The two most widespread and often used VSC variants are the three leg VSC for three-phase three wire (TPTW) and the four leg VSC for three-phase four wire (TPFW). Conventional control algorithm SRFT (Synchronous Reference Frame Theory) and two not so conventional Notch Filter and Adaptive Notch Filtering distribution systems have been simulated, developed and analyzed in this work. For power systems to operate safely and reliably, mitigation of these PQ problems is necessary and a lot of literature is published and is being published presenting various techniques for PQ problem mitigation. The administration and operation of modern power systems have benefited greatly from SAPF. The most current research initiatives and solutions strive to design a suitable compensator for quick and efficient compensation. Before implementing hardware, system design involves mathematical modelling and analysis. In their respective papers, the reasons for voltage sags in industrial plants are discussed by Mark F. McGranaghan et al [1], emphasizing that system faults are the main cause of voltage sags. P. Wang et al. [2] focuses on the investigation of voltage sags and the accompanying phase angle jump. Modelling and performance analysis of a shunt compensated system was demonstrated by D. Kumar and Rajesh [3] where they considered unbalanced linear and nonlinear loads.

For optimal performance of shunt compensation filters, the control algorithms are key factors. It is commonly understood that the controller's capabilities and speed determine how well the SAPF unit function when faced with an unpredictable load. and speed to provide reference current for compensation. The converter's insulated gate bipolar transistors (IGBTs) receive PWM pulses from the control algorithm via a digital method. Numerous control methods have been published in the literature to accomplish this. These tend to be split into time-domain algorithms and frequency-domain algorithms.

D-Q reference frame theory [4], Instantaneous Reactive Power Theory (IRPT) [5], back propagation algorithm [6], etc. are some of the well-known algorithms. P. Chittora, A. Singh and M. Singh [7] presents a notch filtering technique of harmonic current extraction and compensation for a TPTW system using a VSC. Simply put, a band stop filter with a notch is one that only lets specific frequencies through. For the extraction of harmonic and reactive current in active power filters, authors in [8] adopted an adaptive notch filtering method. The primary goal of this technology

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:35 UTC from IEEE Xplore. Restrictions apply.

is to deliver synchronized reactive and harmonic current components for control.

Here, three distinct methods of generating reference current and extracting reactive and harmonic current components are discussed. SRFT, i.e., Synchronous Reference Frame Theory is one of the effective and straightforward ways of delivering reference currents. Notch Filtering is also presented to calculate the load current and voltage's in phase and quadrature components. Further, adaptive notch is also used which can effectively identify and monitor alterations in the observed signal's frequency and isolate the time-varying harmonics. Then, using a PWM current controller, the switching pulses for the 6 IGBTs in the VSC bridge are determined through comparison of the reference currents with the actual currents for 3-ph VSC. We are looking at the VSC that is employed as the DSTATCOM (Distribution Static Compensator) in power factor correction mode. Simulation results are also presented.

II. Description of the Grid System

Figure 1 showcases the structure of the system model. The major elements of the modelled system are voltage supply from grid, three phase three wire (TPTW) VSC acting as SAPF, linear & nonlinear loads. Three phase, 110V line-to-line, 50 Hz grid voltage is supplied. The system consists a R-L nonlinear load connected in parallel with the diode rectifier at last. The supply-compensator-load are linked up at the PCC. The SAPF is represented by a VSC (TPTW configuration) linked with a shunted DC link capacitor. It is also the duty of the DC link capacitor to generate adequate compensatory electrical currents. 6 gate pulses are mandatory for VSC to operate properly.

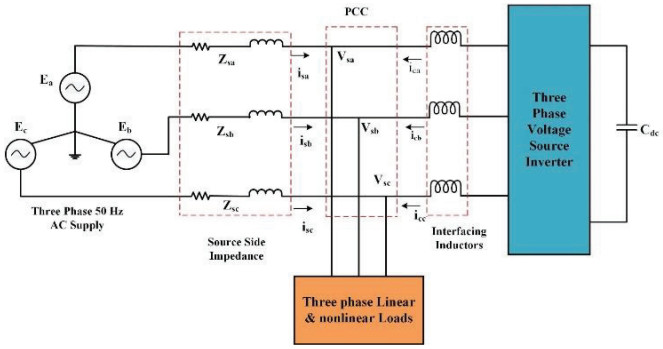


Fig. 1. TPTW system which is being considered.

III. DSTATCOM Control Algorithm

A. SRFT - Synchronous Reference Frame Theory

One simplistic and most common technique of generating reference current is SRFT which is a time domain-based approach. The 3-phase load current in an abc stationary frame is converted into direct and quadrature axis components making it easier for a LPF (low pass filter) to minimize harmonic aspects. SRFT is used to extract fundamental active power component of the load current [9]. For calculating stationary load current components, the aforementioned technique relies on Clarke and Park transformation of load current.

1) Clarke Transformation

$$\begin{bmatrix} i_{L\alpha} \\ i_{L\beta} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix} \begin{bmatrix} i_{La} \\ i_{Lb} \\ i_{Lc} \end{bmatrix} \quad (1)$$

2) Park Transformation

$$\begin{bmatrix} I_{Ld} \\ I_{Lq} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} Cos\,\theta & Sin\,\theta \\ -Sin\,\theta & Cos\,\theta \end{bmatrix} \begin{bmatrix} i_{L\alpha} \\ i_{L\beta} \end{bmatrix} \quad (2)$$

SRFT control block diagram is depicted in the figure 2. The process of converting AC quantities (abc) into DC quantities (dq) done in 2 steps: first involves Clarke's transformation shown in equation 1 and second involves Park transformation as given in eq 2. I_{Ld} and I_{Lq} are provided by the above transformations and I_{Ld} component is passed into an LPF.

The PI controller gives the source current's loss component (I_{loss}) which when added to the load current (which is DC after abc to dq0 transformation) gives the dq frame reference current (I_{effd}). The PI controller is expressed in the below equation:

$$G(s) = K_p + K_i/s$$

here K_p = proportional gain constant, K_i = integral gain constant.

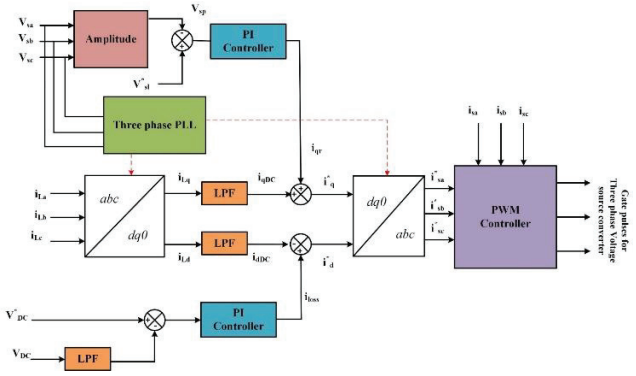


Fig. 2. SRFT Control Algorithm

The AC quantities are obtained back by applying Reverse Clarke's and Park's conversions shown in equations 3 and 4 respectively.

$$\begin{bmatrix} i^*_{s\alpha} \\ i^*_{s\beta} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} Cos\,\theta & -Sin\,\theta \\ Sin\,\theta & Cos\,\theta \end{bmatrix} \begin{bmatrix} I_{effd} \\ 0 \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} i^*_{sa} \\ i^*_{sb} \\ i^*_{sc} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} 1 & 0 \\ -1/2 & \sqrt{3}/2 \\ -1/2 & -\sqrt{3}/2 \end{bmatrix} \begin{bmatrix} i^*_{s\alpha} \\ i^*_{s\beta} \end{bmatrix} \quad (4)$$

B. Notch Filter Based Control Technique

After tuning itself to a certain frequency, the Notch Filter determines the value of the tuned frequency constituent. It at best permits one frequency portion signal to pass and inhibits any other unwanted additional frequencies. This NF is used to fetch the supply system's fundamental frequency (50 Hz) component, which is then modulated to improve PQ in the 3 phase 3 wire TPTW distribution with linear as well as nonlinear loads.

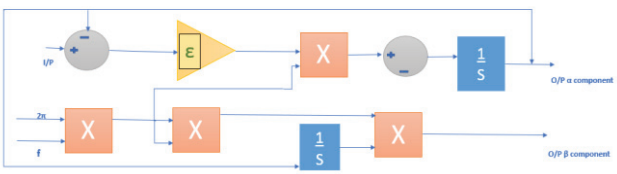


Fig. 3. Notch Filter block diagram

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:35 UTC from IEEE Xplore. Restrictions apply.

Figure 3 depicts a block depiction diagram of notch filter algorithm. It is used to figure out signal's in phase (α) & quadrature phase (β) components. The transfer function is expressed as

$$G(s,\emptyset) = \frac{s^2 + \emptyset^2}{s^2 + 2s\varepsilon\emptyset + \emptyset^2} \quad (4)$$

Here, ε = damping constant and $\emptyset$ = phase angle. For 50Hz system, phase angle $\emptyset = 2\pi f => 2 \text{ x } \pi \text{ x } 50 => 314\ rad/s$. Given below are the expressions involving the relationship between settling time (Ts), damping constant (ε) and bandwidth ($\Delta\omega$):

$$\Delta\omega = 2 * \omega_0 * \varepsilon \quad (5)$$

$$Ts = -\frac{ln(\vartheta)}{\varepsilon * \omega_n} \quad (6)$$

Here, ϑ = tolerance fraction 2%, ω_0 = fundamental frequency, ω_n = natural frequency of oscillation. From eq 5 and 6, we can see that whereas the settling time is inversely correlated with the damping constant, the bandwidth is directly correlated with it. Thus, if we decrease value of damping constant, bandwidth also decreases which we want as it leads to high quality factor but we see an increase in the settling time which we don't want. Thus, keeping both the possibilities in mind we have chosen ε to be 0.45.

The Simulink model in figure 4 shows the production of gate pulses for VSC. The fundamental component of load current 'I_{Laest}', 'I_{Lbest}', 'I_{Lcest}' for phase a-b-c respectively obtained using notch filtering are averaged to obtain I_{avg} . This I_{avg} is helpful to measure the value of the referencing currents explained later.

The losses happening inside VSC are represented by the PI controller's output (i_{dc}). If the supply supplies the in-phase load current and covers the expense of losses, then the weight update eq is given by,

$$w_{pt} = i_{dc} + w_{pavg} \quad (7)$$

where w_{pavg} = weight avg of 3-ph load current.

The weight gained (w_{pt}) is multiplied by in phase unit template u_{pa}, u_{pb}, and u_{pc} provided in Eq. 8 to produce three phase supply reference currents.

$$u_{pa} = \frac{v_{ta}}{v_t};\ u_{pb} = \frac{v_{tb}}{v_t};\ u_{pc} = \frac{v_{tc}}{v_t} \quad (8)$$

$$i_{sa}^* = i_{L\alpha} \text{ x } u_{pa},\ i_{sb}^* = i_{L\alpha} \text{ x } u_{pb},\ i_{sc}^* = i_{L\alpha} \text{ x } u_{pc} \quad (9)$$

Thus, the reference currents isa*, isb*, isc* fetched as shown in equation 9.

C. *Adaptive Notch Filter*

Figure 5 portrays the typical implementation of an adaptive notch filter using a block diagram used to detect load current. The transfer function of adaptive notch is described as [10]:

$$T(s,\emptyset) = \frac{s^2 + \theta^2}{s^2 + \zeta\theta + \theta^2} \quad (4)$$

The frequency is considered constant. This filter has ability to control ζ. The computational efforts in the notch is decreased due to constant frequency assumption. In essence, the notch filter separates the essential part from the input signal using two integrators. By altering the damping constant, bandwidth and settling time can be changed.

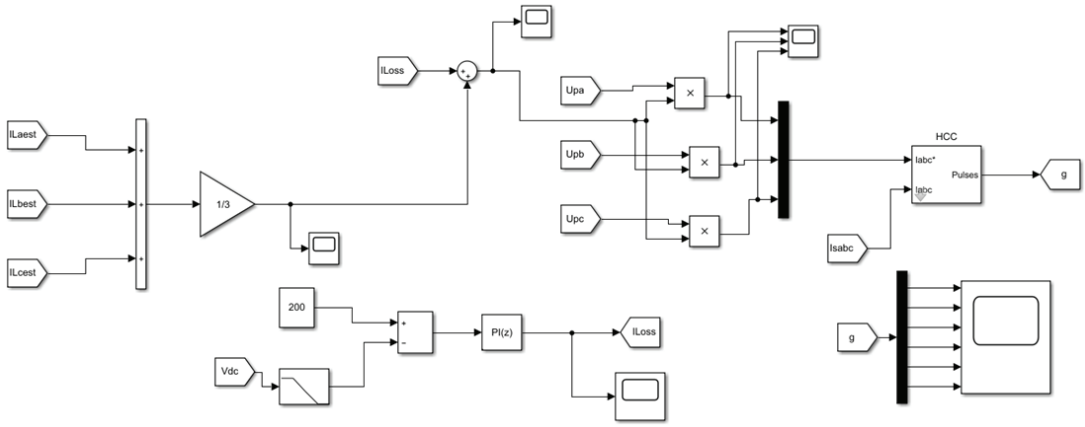


Fig. 4. VSC Gate pulse generation

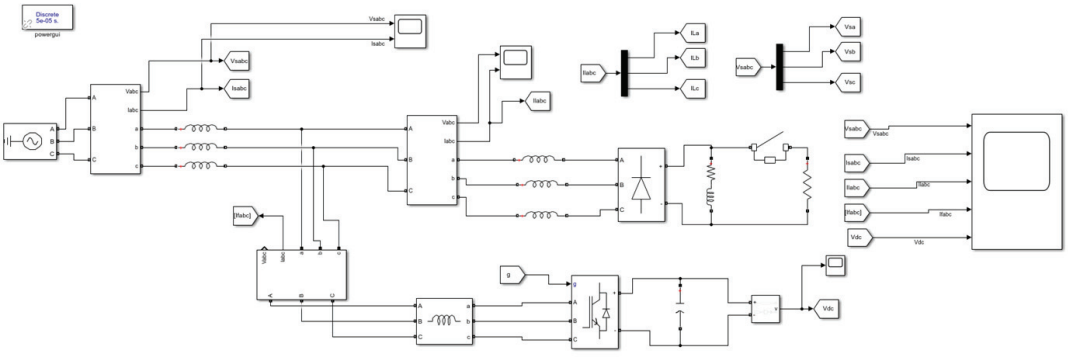


Fig. 5. System's modelling in Simulink

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:35 UTC from IEEE Xplore. Restrictions apply.

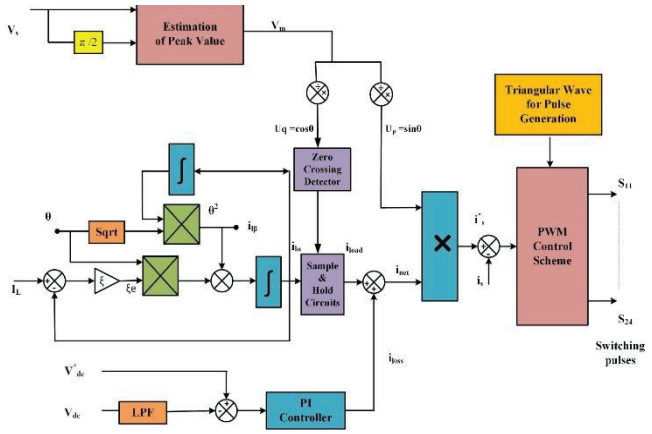


Fig. 6. Adaptive Notch Filter block diagram

IV. Results and Discussions

This section covers the steady state waveforms and simulation findings. Fig 6-8 portrays the steady state waveforms of V_{sabc} source voltage, I_{sabc} source current, I_{Labc} load current, I_{invabc} injected current (compensating current from DSTATCOM) and V_{dc} the dc link voltage under simulation of SRFT, Notch and Adaptive Notch control algorithms respectively. A diode rectifier is connected in system to portray a nonlinear load with R = 7.5 Ω resistance, L = 40 * 10^{-3} H inductance. The VSC DC voltage is controlled to 200 V with a PI controller. There is an imbalance on the load side as well. In Fig 9-11, system performance is examined. As shown in the figures are balanced supply currents, sinusoidal in steady and dynamic system operation.

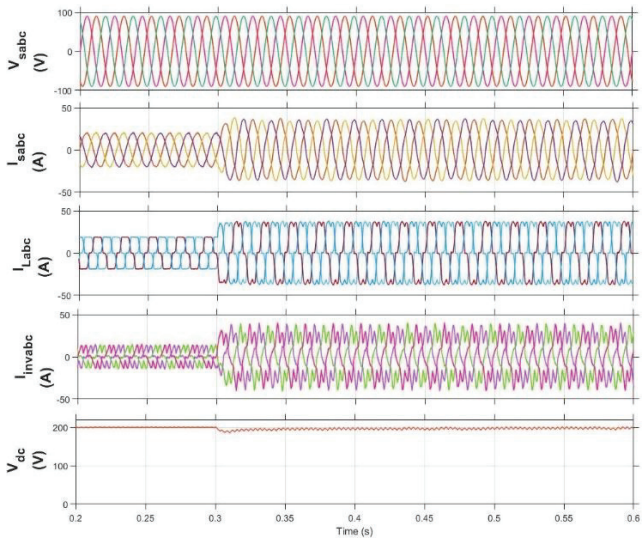


Fig. 7. SRFT Simulation waveform results obtained

TABLE I. PARAMETER OF THE SYSTEM

Source RMS Voltage	110V
Inductance (Source)	0.003H
Inductance (Load)	0.002H
Load Resistance	7.5Ω
Interfacing Inductors	0.002H

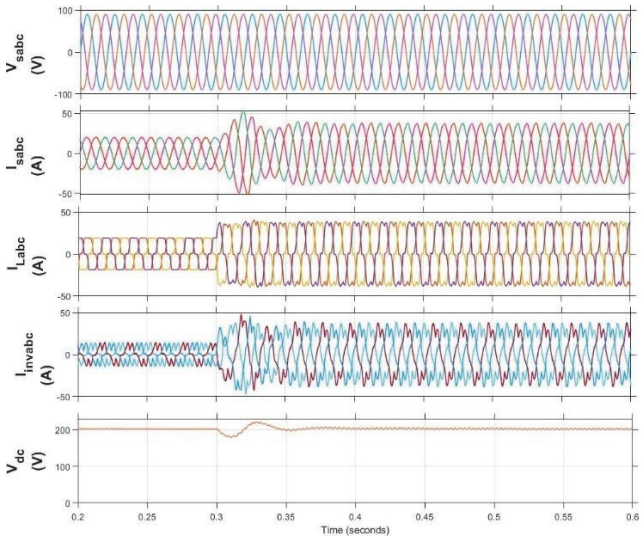


Fig. 8. Notch Filter Simulation waveform results obtained

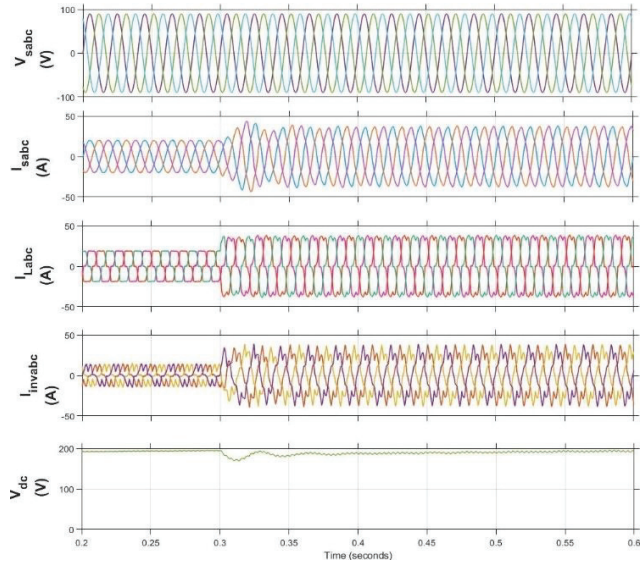


Fig. 9. Adaptive Notch Filter Simulation waveform results obtained

In each of the three control algorithm implementations, the source current can be seen to be almost in phase with the supply voltage. Thus, it can be said that these techniques are employed not only to compensate for the load but also capable of improving the power factor to 1. Also, it can be seen during the instant t = 0.3s in Fig 6-8, as the load is increased, the DC connection voltage briefly drops, but it quickly returns to its constant value of 200V because of the self-regulation provided by the PI controller. The source current increases to meet the demand when the load is increased at 0.3 s, but it preserves its sinusoidal shape and continues to follow the in-phase grid voltage. However, all three of the implementations show prolonged oscillations.

In accordance with IEEE 519 standards, it can be seen from Fig. 10-12 that THD of I_{sabc} source current decreased below the range of 5% with the use of SRFT, NF, and ANF. At t = 0.2 s when the load is increased, the dc link capacitor voltage waveform stabilizes in less than 0.04 seconds displaying minor disruption at t = 0.3s when the load is increased before quickly settling.

The effectiveness of the control algorithms in lowering the THD of I_{sabc} source current is demonstrated in Fig. 10–

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:35 UTC from IEEE Xplore. Restrictions apply.

12. As can be seen in the figures 7-9, the source current is essentially sinusoidal. Under SRFT, Notch Filter, and ANF, the THD (Total Harmonic Distortion) of the source current is minimized to 4.92%, 2.38%, and 1.92%, respectively.

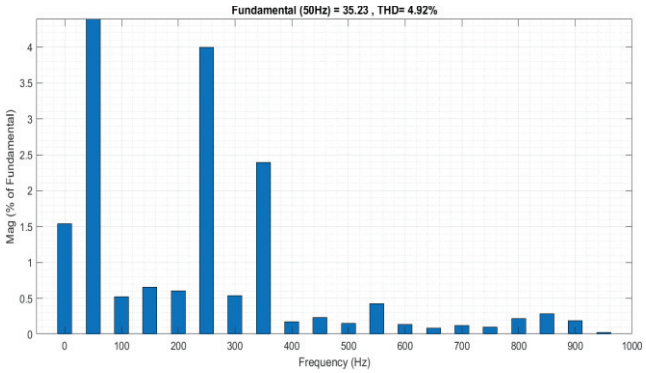


Fig. 10. I_{sabc} Source Current THD obtained by SRFT

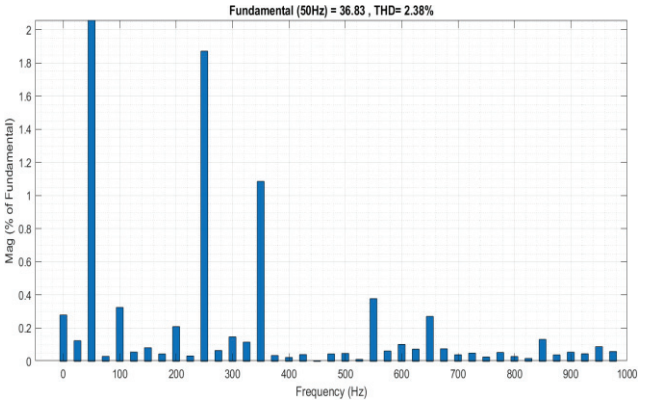


Fig. 11. I_{sabc} Source Current THD obtained by Notch Filter

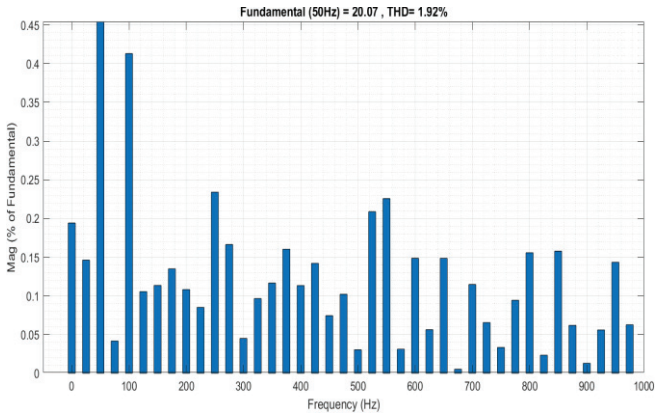


Fig. 12. I_{sabc} Source Current THD obtained by ANF

TABLE I. COMPARISON TABLE OF PERFORMANCE

Control Algorithm	*Oscillations*	*Peak Overshoot (V)*	*Settling Time (s)*	*% THD (I_{sabc})*
SRFT	*High Level Oscillations*	*2.8*	*0.06*	*4.92*
Notch	*Medium Level Oscillations*	*1.13*	*0.05*	*2.38*
ANF	*Negligible Oscillations*	*0.44*	*0.04*	*1.92*

Figure 13 displays a comparative study of the fundamental current components extracted of each of the control algorithms. The settling time for SRFT is 0.06 seconds, for notch filtering it is 0.05 seconds, and for adaptive notch filtering it is 0.04 seconds. The peak overshoot obtained for both SRFT and NF is 2.8V and 1.13V with continuous oscillations in steady state as well as dynamic, and for ANF it's 0.44V with barely perceptible oscillations.

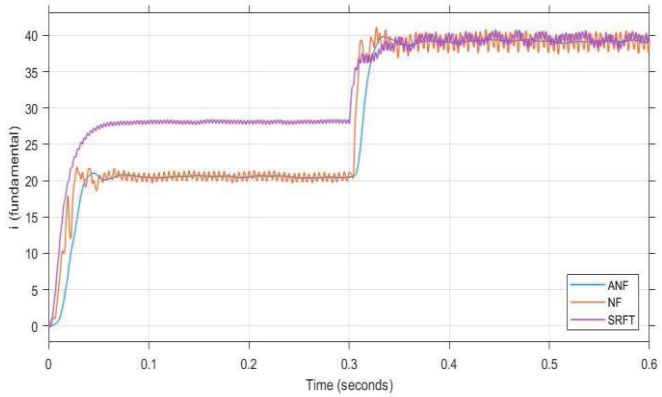


Fig. 13. Fundamental current component extracted using SRFT, NF and ANF

V. CONCLUSIONS

This paper illustrates the comparative analysis of SRFT (Synchronous Reference Frame Theory), NF (Notch Filter) and ANF (Adaptive Notch Filter) control methodology. The algorithms are successfully modeled in Simulink and the resulting observations are presented. Each one's dc side capacitor voltage and fundamental current component are contrasted. Each methodology's overall harmonic distortion is examined. The resulting FFT analysis illustrates that THD of source current reduced in SRFT is found to be 4.92%, 2.38% in NF and 1.92% in ANF.

One of the primary features of the adaptive notch configuration is that minimal THD (1.92%) is achieved even at a relatively low load current magnitude. The adaptable characteristics of the adaptive notch filter additionally allows the system to function better in unpredictable load scenarios. It has ideal tracking capabilities and reacts fast to load changes. Finally, it can be reported that all of the control approaches put into place have been successful in preventing the grid supply's power quality from degrading, and the final THD of I_s source current is way below threshold set by IEEE 519.

REFERENCES

[1] M. F. McGranaghan, D. R. Mueller and M. J. Samotyj, "Voltage sags in industrial systems," in IEEE Transactions on Industry Applications, vol. 29, no. 2, pp. 397-403, Mar/Apr 1993.doi: 10.1109/28.216550.

[2] P. Wang, N. Jenkins and M. H. J. Bollen, "Experimental investigation of voltage sag mitigation by an advanced static VAR compensator," in IEEE Transactions on Power Delivery, vol. 13, no. 4, pp. 1461-1467, Oct 1998.doi: 10.1109/61.714772.

[3] D. Kumar and Rajesh, "Modelling, Analysis and Performance of a DSTATCOM for Unbalanced and Non-Linear Load," 2005 IEEE/PES Transmission & Distribution Conference & Exposition: Asia and Pacific, Dalian, 2005, pp. 1-6. doi: 10.1109/TDC.2005.1547142.

[4] Gonzalez-Espin, F., Figueres, E., Garcera, G.: "An adaptive synchronous reference-frame phase-locked loop for power quality improvement in a polluted utility grid', IEEE Trans. Ind. Electron., 2012, 59, (6), pp. 2718–2731

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:35 UTC from IEEE Xplore. Restrictions apply.

[5] Herrera, R.S., Salmerón, P., Kim, H.: “Instantaneous reactive power theory applied to active power filter compensation: different approaches, assessment, and experimental results’, IEEE Trans. Ind. Electron., 2008, 55, (1), pp. 184–196

[6] B. Singh and S. Arya, “Back-propagation control algorithm for power quality improvement using DSTATCOM,” IEEE Trans. Ind. Electron., vol. 61, no. 3, pp.1204-1212, Mar. 2014

[7] P. Chittora, A. Singh and M. Singh, "Harmonic current extraction and compensation in three phase three wire system using notch filter," 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum, India, 2015, pp. 422-427, doi: 10.1109/RAICS.2015.7488453.

[8] D. Yazdani, A. Bakhshai, G. Joos, M. Mojiri,” An adaptive notch filtering approach for harmonic and reactive current extraction in active power filters”, Proceedings - 34th Annual Conference IECON 2008, pp. 535–538. 2008.

[9] M. Labeeb and B. S. Lathika, "Design and analysis of DSTATCOM using SRFT and ANN-fuzzy based control for power quality improvement," 2011 IEEE Recent Advances in Intelligent Computational Systems, Trivandrum, 2011, pp. 274-279. doi: 10.1109/RAICS.2011.6069317.

[10] P. Bansal and A. Singh, "Adaptive Notch Filter Control Algorithm for 5-Level Distribution Static Compensator," 2020 IEEE First International Conference on Smart Technologies for Power, Energy and Control (STPEC), Nagpur, India, 2020, pp. 1-5, doi: 10.1109/STPEC49749.2020.9297721.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:22:35 UTC from IEEE Xplore. Restrictions apply.

Reversible Data Hiding scheme for color images based on skewed histograms and cross-channel correlation

Priyansh Bhatnagar
Delhi Technological University
Delhi, India
priyanshbhatnagar2002@gmail.com

Prateek Tomar
Delhi Technological University
Delhi, India
tomarprateek09@gmail.com

Rishabh Naagar
Delhi Technological University
Delhi, India
rishabhnaagar1502@gmail.com

Rajeev Kumar
Delhi Technological University
Delhi, India
rajeevkumar@dtu.ac.in

2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT) | 979-8-3503-3936-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/APSIT58554.2023.10201792

***Abstract*—Reversible data hiding (RDH) is an undetectable communication technology that enables restoration of the original cover image after extracting the hidden secret data from the image. It's being widely used in sevaral fields such as military, medical, etc,. Till date, various RDH have been developed to increase the amount of hidden data while maintaining their quality. However, the problem of traditional trade-off between the embedding capacity (EC) and image quality still persists. For this, we propose a new RDH method for color images based on skewed histograms and cross-channel correlation. The skewed histogram is generated with the help of extreme predictors, aids in achieving less distortion by only incorporating pixels from the peak and short tail. Further, the proposed method splits the secret data (or payload) based on each channel's characteristics so that the most smooth channel can be exploited for embedding to further benefit in reducing the caused distortion and increasing the EC. Moreover, embedding in the complex regions of images is done with the help of more comprehensive predictor which also takes into account various possible edges. Thus, the proposed method achieves greater EC with better quality stego-images than the related and existing RDH methods for color images.**

***Index Terms*—RDH, Color image, Skewed Histogram, Edge Detection, Extreme Predictors**

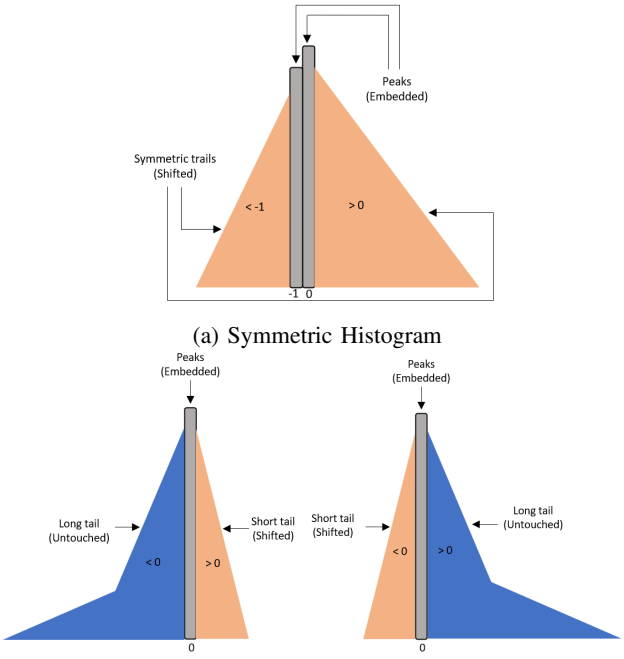
I. Introduction

Reversible data hiding provides an efficient method for storing information in an image, while still allowing the original image to be retrieved when needed [1]. There are various applications for reversible data hiding, including distortion-sensitive uses such as military and medical imaging. In these cases, authentication data is embedded as a watermark, and image enhancement filters can be applied directly to the watermarked image, while the original image remains intact and can be recovered as necessary [2]. Furthermore, data can be extracted from the watermarked image as required, enabling secure and efficient information storage. Reversible data hiding is also applied where the goal is to achieve contrast improvement, reducing file size and obscuring data simultaneously.

In order to achieve reversible data hiding for applications that are sensitive to distortion, various techniques have been developed. Researchers have proposed Reversible data hiding algorithms that utilize lossless image compression and histogram shifting to modify histograms while hiding the data. Modern RDH algorithms frequently use prediction-error expansion (PEE) and histogram shifting methods as their foundation. Histogram shifting is a method for hiding data that includes shifting a group of histogram peaks in a reversible approach utilising a pixel histogram or a modified version of the pixels. There are other variations of this approach, including difference expansion, between a pixel and its nearest neighbour, and pixel-based histogram shifting, which use the most common pixels for data Hiding. The use of histogram shifting in prediction error histograms which is widely known as Prediction error expansion (PEE), as opposed to the use of pixel or pixel difference histograms, is more advanced and requires less adjustment to hide data. Moreover, it has been shown to be effective to embed the algorithm into the smooth portions of the host image and with many passes.

Contrary to its rapid advancement for gray-scale images, the development of RDH (Reversible Data Hiding) for color images has not received enough attention. Even though color images are more frequently used in daily life, little research has been done on RDH methods to process color images [3]. In this paper, a new reversible data hiding method for color images is proposed, which uses a more comprehensive predictor than existing methods and simultaneously exploits cross-channel relation (Edge information from reference color channels and adaptive partitioning) present in color images and skewed histogram shifting with an extended sorting method to efficiently embed payload data into RGB channels.

979-8-3503-3936-9/23/$31.00 ©2023 IEEE
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:57 UTC from IEEE Xplore. Restrictions apply.



(a) Symmetric Histogram

(b) Left Skewed Histogram (c) Right Skewed Histogram

Fig. 1: Symmetric vs Skewed Histograms [4]

an overview of previous RDH methods based on Prediction Error Expansion (PEE) used for gray-scale images and several existing RDH schemes for color images. In Section III, we describe the implementation of the proposed method. Section IV presents the experimental results and analysis. Finally, Section V provides the conclusion of the paper.

II. Related Work

In this section some of the existing and related RDH schemes for color images are briefly reviewed.

It has been observed that the focus of current RDH research has been mainly on grayscale images [5]–[8], despite the popularity of color images in reality. Only a handful of RDH methods exist for color images. In 2004, Alattar [3] introduced an RDH method based on efficient integer transformation. This method independently embeds the secret data into each color channel and found to be more effective than applying it across the channels. Tang *et al.* [9] used a double-layer embedding technique based on HS to improve capacity, while Qu *et al.* [10] proposed a greedy algorithm for partition parameters. Yang *et al.* [11] proposed an RDH method in color filter array images, which are raw signals from digital cameras, is unsuitable for real color images. Zhou *et al.* [12] modify RGB channels to approximate a rate-distortion model which is modification-independent in order to achieve in-variance and Hou *et al.* [13] maintained gray-scale in-variance.

To enhance the performance of reversible data hiding methods for color images by utilizing inter-channel correlation to exploit redundancy present in color images, two strategies were created [14], [15] where Li *et al.* [14] focused on improving prediction performance while Ou *et al.* [15] focused on payload partition. In the former, edges in the reference channel were used to predict pixels in the current channel, utilizing edge direction in edge areas and rhombus prediction in smooth areas. This approach has drawbacks, though, and can be improved. In the latter, the best candidate with the shortest sub-sequence length was selected after payload sizes were adaptively given to each channel. To further enhance the performance, the proposed method incorporates both strategies, improving prediction performance with edge information and adaptive payload splitting and further reducing distortion using skewed histograms.

III. Proposed Work

As discussed in the previous section, the distortion in gray scale images while data hiding can be further reduced by using skewed prediction error histogram instead of symmetric prediction error histogram to embed payload [5]. The goal is to move pixels in short tail of skewed histogram and keep long tail mostly unchanged as by Kim et al. [4]. While using edge information for prediction and adaptive partitioning of payload for RGB channels, the proposed method can embed data into color images by calculating prediction errors of each channel and embedding data into channel separately and the distortion generated in the watermarked image will be significantly lower than by using symmetric prediction error histogram method. For this, a pair of extreme predictions is used i.e., a high estimate and a low estimate. High estimate are used to make positive histogram shifting (PHS) while low estimate make negative histogram shifting (NHS). It is to be noted that NHS only permits histogram bins to move inside the negative direction while PHS only permits histogram bins to move inside the positive direction. Also, to exploit inter-channel correlation in color images, edge information is used from reference color channel so that the pair of extreme predictions are calculated differently for smooth and edge pixels to improve the prediction performance and increase embedding capacity. Moreover, to further reduce embedding distortion, instead of embedding payload equally into each color channels, a comprehensive search is performed by calculating embedding distortions to find optimal payload partitioning for each color channels.

The discussion on the proposed work is divided into 9 subsections where the first section i.e., III-A discusses about how to find pair of extreme predictions. In section III-B these predictions are made more comprehensive using edge information from reference channels. Sections III-C and III-D show how embedding and extraction are done by PHS and NHS. Section III-E shows how local complexities are going to be calculated which are used for sorting and adaptive partitioning which is discussed in III-F. Section III-G discusses about pre-processing of the image before encoding and lastly we discuss about encoding and decoding algorithms in sections III-H and III-I respectively.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:57 UTC from IEEE Xplore. Restrictions apply.

A. Pair of Extreme Predictions

In traditional rhombus prediction method, all of the nearest neighbors of the current pixel are used to find prediction value. In order to perform skewed histogram shifting, neighboring pixels' values (n, w, s, e) are arranged in ascending sequence, as in $x_1 \leq x_2 \leq x_3 \leq x_4$. While Predictor 3 uses the majority of neighbours to forecast the pixel value, Predictor 1 uses the highest and closest neighbouring values. Prediction value is calculated as shown in Table-I. Since double-

TABLE I: Predictor pairs. [.] rounds to the nearest integer.

Predictor No.	1	2	3
$\hat{p}_h$	x_4	$\left[\frac{x_4+x_3}{2}\right]$	$\left[\frac{x_4+x_3+x_2}{3}\right]$
$\hat{p}_l$	x_1	$\left[\frac{x_1+x_2}{2}\right]$	$\left[\frac{x_1+x_2+x_3}{3}\right]$

embedding around 0 is known to be ineffective, this method does not analyse embedding using the same prediction values. This is avoided by using the equation below to guarantee that the pair of prediction numbers is always distinct.

$$\hat{p}_l = \begin{cases} \hat{p}_l - 1 & \text{if } \hat{p}_h = \hat{p}_l \\ \hat{p}_l & \text{otherwise} \end{cases} \tag{1}$$

B. Edge Detection in reference channel

Each color channel's data is hidden sequentially, but during the prediction phase, for some samples in the current channel, the accuracy of the prediction is increased by using edge information from a different channel, known as the reference channel. In this method, horizontal and vertical edges are considered and pixels lying on these edges will be predicted accordingly. Distance between the nearest neighbors and the current pixel is determined by value D_m, which is given by

$$D_m = \left| \sum_{k=1}^{4} \gamma_m^k I_r^k - I_r \right| \tag{2}$$

where $\gamma_m^k = \frac{1}{4},\ \frac{1}{4},\ \frac{1}{4},\ \frac{1}{4}$ are coefficients for $k = 1,\ 2,\ 3,\ 4$ considering 1, 2, 3, 4 to be n, w, s, e neighbors respectively, and I_r^k represent four nearest neighbours of a pixel in the reference channel and I_r represents that pixel in reference channel. Next, the edge direction is calculated(to check if it is a horizontal edge or a vertical edge). This is estimated by calculating D_h and D_v values.

$$D_h = \left| \frac{I_r^w + I_r^e}{2} - I_r \right|, \tag{3}$$

$$D_v = \left| \frac{I_r^n + I_r^s}{2} - I_r \right| \tag{4}$$

Minimum of these values is considered to be D_e, which determines the edge direction.

$$D_e = min\left\{D_h,\ D_v\right\}. \tag{5}$$

$$D_e = \left| \sum_{k=1}^{4} \gamma_e^k I_r^k - I_r \right| \tag{6}$$

where γ_e^k are coefficients which are decided according to edge direction. For example, if it is horizontal direction $\gamma_e^k = 0,\ \frac{1}{2},\ 0,\ \frac{1}{2}$ for $k = 1, 2, 3, 4$.

When examining an image, it can be seen that smooth areas have pixels with values that are fairly similar, leading to small values for D_m and D_e as well as a slight variation between them. A smaller value for D_e but a higher value for D_m results from differences in the neighbouring pixels in edge regions, where pixels are only similar to those that face the same edge direction. By computing $(D_m - D_e)$, it is possible to identify if the reference sample is in an edge region. This technique can also be used to categorize the location of the current sample because the edge distribution of all colour channels is similar. The current sample is considered to be located at or near an image edge if $(D_m - D_e)$ is greater than a predetermined threshold value (τ) and the prediction process should take edge information from another channel into consideration. On the other hand, using edge information wouldn't help much if $(D_m - D_e)$ is less than (τ). The threshold value (τ) is chosen from a predetermined range $(0, 1, ..., 10)$ based on which embedding rate results in the highest PSNR value.

If $(D_m - D_e)$ is greater than the threshold value (τ), it is likely that the current sample is located at an edge area, and the prediction value is calculated by utilizing edge information estimated from the reference channel, which is given by:

$$\hat{p}_h = max\ (2\gamma_e^k I_c^k,\ k \in 1,\ 2,\ 3,\ 4) \tag{7}$$

$$\hat{p}_l = min\ (2\gamma_e^k I_c^k,\ k \in 1,\ 2,\ 3,\ 4) \tag{8}$$

Where γ_e^k represent the coefficients used according to edge direction and I_c^k represents neighbouring pixels values in current channel.

If $(D_m - D_e)$ is smaller than or equal to the threshold value (τ), incorporating edge information is unnecessary, and the prediction value is obtained simply by Predictor 1, 2, 3 as discussed in Section III-A.

C. Embedding

Two time embedding is done in every pixel, once using the high estimate $\hat{p}_h$ and once using the low estimate $\hat{p}_l$. The difference between the pixel value (p) and the high estimate $(\hat{p}_h)$ creates the initial skewed prediction error histogram. In this step, PHS is used to embed one bit in the pixel value (p). (skewed histogram shifting based on the high estimate).

$$p^{'} = \begin{cases} p + m_1 & \text{if } p - \hat{p}_h = 0 \\ p + 1 & \text{if } p - \hat{p}_h > 0 \\ p & \text{otherwise} \end{cases} \tag{9}$$

Only when the prediction error is 0 is one bit embedded. Other pixels' prediction errors are shifted by 1 for them. Using PHS, this is accomplished in the first step. Creating another skewed prediction error histogram in the second step entails calculating the difference between the pixel value $(p^{'})$ and the low estimate $(\hat{p}_l)$. During this step, one bit is embedded in the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:57 UTC from IEEE Xplore. Restrictions apply.

(a) Red (b) Green (c) Blue

Fig. 2: Canny Edge Detection performed on different color channels in Airplane image

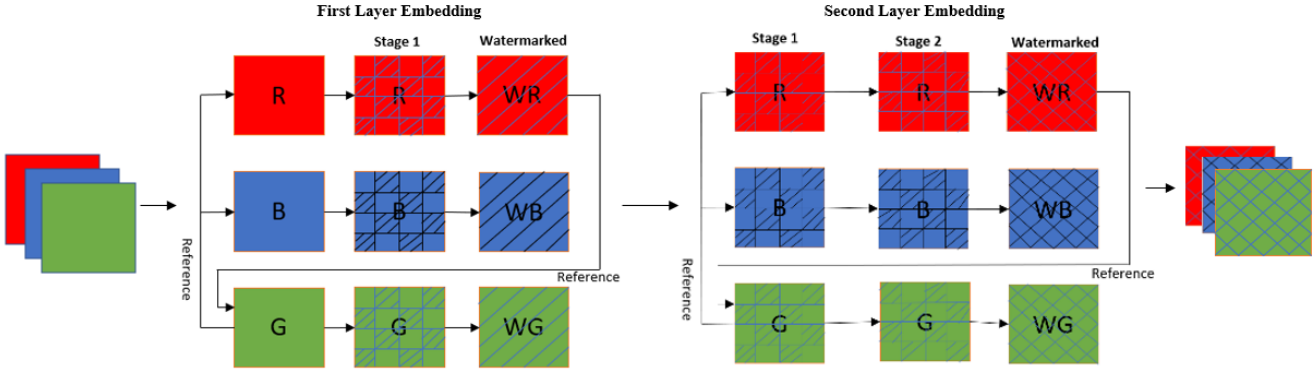


Fig. 3: Schematic Diagram for Double layer embedding with edge information from reference channels

pixel value $(p^{'})$ using NHS (skewed histogram shifting based on the low estimate).

$$p^{''} = \begin{cases} p^{'} - m_2 & \text{if } p^{'} - \hat{p}_l = 0 \\ p^{'} - 1 & \text{if } p^{'} - \hat{p}_l < 0 \\ p^{'} & \text{otherwise} \end{cases} \tag{10}$$

During the second step using NHS, one bit is embedded only when prediction error is zero. For other pixels, their prediction errors are shifted by 1.

D. Extraction

The steps that were taken during embedding are reversed throughout the extraction and recovery phase. The even pixels are restored following the odd pixels, according to this. The odd pixels are sorted first. Also, the payload is extracted from the pixel in the reverse order i.e., extract using NHS (by using $\hat{p}_l$) first then by PHS (by using $\hat{p}_h$). Payload and first watermarked pixel are recovered using NHS as follows:

$$p^{'} = \begin{cases} p^{''} + 1 & \text{if } p^{''} - \hat{p}_l < 0 \\ p^{''} & \text{otherwise} \end{cases} \tag{11}$$

$$m_2 = \begin{cases} 0 & \text{if } p^{''} - \hat{p}_l = 0 \\ 1 & \text{if } p^{''} - \hat{p}_l = 1 \end{cases} \tag{12}$$

Now the original pixel and second payload bit are recovered using PHS as follows:

$$p = \begin{cases} p^{'} - 1 & \text{if } p^{'} - \hat{p}_h > 0 \\ p^{'} & \text{otherwise} \end{cases} \tag{13}$$

$$m_1 = \begin{cases} 0 & \text{if } p^{'} - \hat{p}_h = 0 \\ 1 & \text{if } p^{'} - \hat{p}_h = 1 \end{cases} \tag{14}$$

E. Local Complexity Functions

The "pixel selection" technique is commonly used in the literature to minimize distortion caused by embedding. This involves ranking pixels based on their local complexity values, which reflect their likelihood of being suitable for embedding. The first pixels to be embedded are those with smaller values because they are thought to be smoother and more embed-able. Two techniques are presented in this subsection: the four-pixel version, and an extended version. Four Pixel Version:

$$C\left(p\right) = \frac{(|N-W|+|W-S|+|S-E|+|E-N|+|N-S|+|W-E|)}{6} \tag{15}$$

Extended Version:

$$\widetilde{C}\left(p\right) = \frac{C(p)+C(NW)+C(NE)+C(SW)+C(SE)}{5} \tag{16}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:57 UTC from IEEE Xplore. Restrictions apply.

F. Adaptive Partitioning

Splitting the payload into three parts according to the RGB channels is a crucial step to minimize distortion during the embedding process in the RGB channels. In this method, the three channels' payloads are assigned to the designations EC_R, EC_G,and EC_B, respectively, and their corresponding embedding distortions are assigned to the symbols ED_R, ED_G, and ED_B. The payload partition optimisation for a given capacity C entails minimising the total distortion across all three channels, which is expressed as:

$$\begin{cases} \text{minimize } ED_R + ED_G + ED_B \\ \text{subject to } EC_R + EC_G + EC_B = C \end{cases} \tag{17}$$

Embedding distortions and embedding capacity are calculated according to the prediction errors calculation in each channel. Below adaptive partitioning method is discussed for single layer embedding.

Let's assume the total embedding capacity of a channel to be C_{max}. For each capacity $\frac{(n \times C_{max})}{1000}$ where $n = 1, 2, 3, \ldots, \frac{C_{max}}{1000}$ iterate through all pixels of a channel and calculate total sum of embedding distortions of pixels and total sum of embedding capacity of all pixels. A maximum bar (t_h) is then put on the local complexity (15) to find out optimal maximum local complexity $(t_h{}^*)$ value a pixel can have to embed data. The idea is to select optimal $t_h{}^*$ values in each channel to get minimum total distortion. In single channel, for each capacity $\frac{(n \times C_{max})}{1000}$, $t_h{}^*$ the one with minimum embedding distortion is found. Embedding distortion for a pixel p and extreme predictors $\hat{p}_h$ and $\hat{p}_l$ is calculated as follows:

$$ed^i{}_1 = \begin{cases} 0.5 & \text{if } p - \hat{p}_h = 0 \\ 1 & \text{if } p - \hat{p}_h > 0 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

$$ed^i{}_2 = \begin{cases} 0.5 & \text{if } p - \hat{p}_l = 0 \\ 1 & \text{if } p - \hat{p}_l < 0 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Where $ed^i{}_1$and $ed^i{}_2$ are embedding distortions by $\hat{p}_h$ and $\hat{p}_l$ respectively. Embedding capacity for a pixel p and extreme predictors $\hat{p}_h$ and $\hat{p}_l$ is calculated as follows:

$$ec^i{}_1 = \begin{cases} 1 & \text{if } p - \hat{p}_h = 0 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

$$ec^i{}_2 = \begin{cases} 1 & \text{if } p - \hat{p}_l = 0 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

Where $ec^i{}_1$ and $ec^i{}_2$ are embedding capacity incremented by $\hat{p}_h$ and $\hat{p}_l$ respectively. These values are calculated for all pixels present in a channel and create vector ED and EC for each channel which contains total embedding distortions and Embedding capacities at each capacity $\frac{(n \times C_{max})}{1000}$. Then iterate through each possible combination of EC_R, EC_G, and EC_B that satisfies the capacity constraint in order to identify the ideal payload partition. Afterwards, find the shortest length of the triple sub-sequence for each combination and select the partition with the minimum length as the optimal solution. The above process is only for single layer embedding, hence it needs to be done for both odd and even halves.

Please note that as this method uses a comprehensive predictor ($\hat{p}_h$ and $\hat{p}_l$ values) which uses edge information also from reference channels, embedding distortion and embedding capacity are calculated in the same manner as the embedding process, i.e., as shown in Fig.3. But as we are finding optimal partition of payload and not embedding data into the channel, the red channel will not be watermarked which is the reference channel for green.So to find accurate adaptive partitioning there is a need to watermark red channel for each capacity and then find optimal partition for only blue and green channels, which in turn increases the algorithm's complexity. Hence for just estimating embedding capacity and distortions, unmodified red channel is used as reference channel for green channel to get an idea of embedding capacity of green channel, and for embedding data modified red channel is used as in Fig. 3.

G. Pre-processing

To avoid issues with underflow and overflow values in the watermarked image, a pre-processing step and location map are necessary. When a pixel is embedded, it might produce a watermarked pixel with values that aren't within the acceptable 0 to 255 range. Performing a pre-processing step on the image prior to embedding can solve this issue by mapping the pixel values of 0 to 1 and 255 to 254. The location map in the payload is subsequently incorporated with the image modification history. The generated location map reads $Ml, ..., Mi, ..., Mn$. The original pixel p_o is pre-processed in the following way to create the pre-processed pixel p:

$$M_i = \begin{cases} 0 & \text{if } p_o = 1 \text{ or } 254 \\ 1 & \text{if } p_o = 0 \text{ or } 255 \end{cases} \tag{22}$$

$$p = \begin{cases} 254 & \text{if } p_o = 255 \\ 1 & \text{if } p_o = 0 \\ p_o & \text{otherwise} \end{cases} \tag{23}$$

To extract the payload embedded in the image, decoder needs auxiliary or side information which will be stored in the LSB's of the image. Auxiliary information to be sent in the LSB's of the image of size $n \times m$ will contain payload length for odd ($t = log_2(n \times m)$ bits) and even ($t = log_2(n \times m)$ bits) halves which will be decided by adaptive partitioning as discussed in section III-F, predictor mode used to embed data (2 bits) and edge threshold value (4 bits) used to determine if a pixel is in edge area or not. Therefore, the total length of auxiliary information will be $2t + 6$ bits.

H. Encoder

According to our proposed method of reversible data hiding in color image, encoding in an image is done as follows:

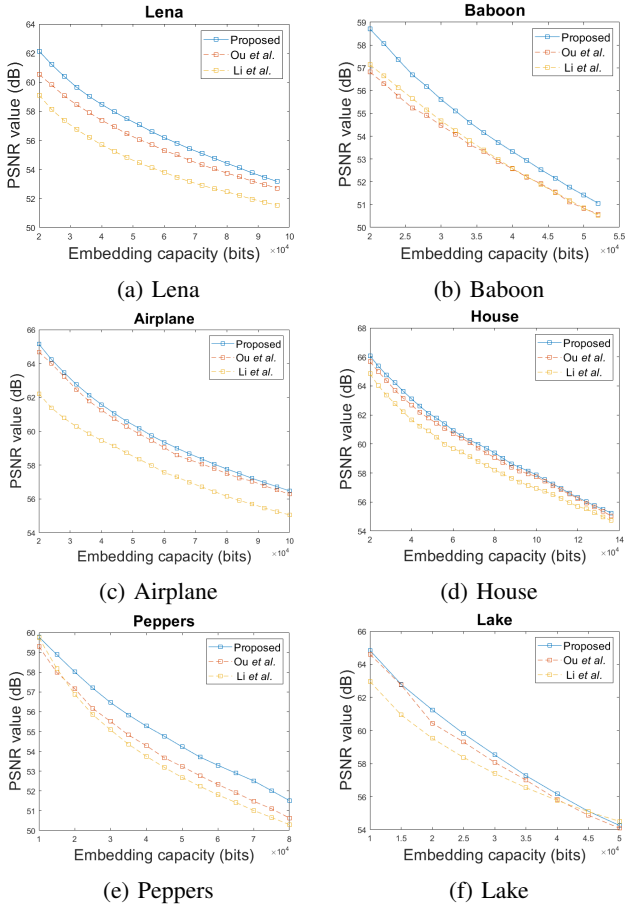
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:57 UTC from IEEE Xplore. Restrictions apply.

1) Decompose the color image of size $n \times m$ pixels into R, G and B channels.
2) Build the location map of each channel and losslessly compress it using arithmetic encoding.
3) The LSBs of $2t + 6$ boundary pixels, length of the compressed location maps and the compressed location map itself are included in the payload.
4) Each channel is represented as a checkerboard where each pixel is classified as even or odd.
5) Find optimal partition of the payload for each channel's even half as discussed in section III-F.
6) Using (16), sort the even pixels.
7) The initial half of the payload is embedded using (9) and (10).
8) Repeat Steps 5-6 for each channel taking reference appropriate reference channels as shown in Fig.3.
9) Repeat Steps 5-8 for embedding in each channel's odd half.
10) The auxiliary information, as discussed in section III-G, is then stored in place of the $2t + 6$ border pixel LSBs of each channel.
11) Combine all marked R, G and B channels to form watermarked image
12) Calculate PSNR by comparing original and watermarked images.
13) Steps 4-14 should be carried out for each predictor mode (1, 2 and 3) and edge threshold (τ).
14) Keep watermarked image with the best PSNR.

I. Decoder

While decoding, we follow reverse of the order we followed in the encoding phase as depicted in Fig.3, i.e., green, blue and then red channel.

1) Watermarked image is decomposed into R, G and B channels.
2) Steps 3-7 will be carried out for each channel in reverse order of encoding (reverse of the order of a single layer embedding shown in fig.-3).
3) The predictor number, payload length for each half and edge threshold are extracted by reading the LSBs of the $2t + 6$ border pixels.
4) Each channel is represented as a checkerboard where each pixel is classified as even or odd.
5) Using (16), sort the odd pixels.
6) To reverse the watermarking process and recover the second payload component in the channel, use (12) and then (13).
7) Using (16), sort the even pixels.
8) To reverse the watermarking process and recover the first component of the payload in the channel, use (12) and then (13).
9) All of the payload components recovered from each channel will be appended to produce complete recovered message
10) For the first $2t + 6$ border pixels, restore the original LSBs for all channels.



(a) Lena (b) Baboon (c) Airplane (d) House (e) Peppers (f) Lake

Fig. 4: Performance Comparison of embedding capacity(bits) with respect to PSNR(dB) on six standard images.

11) Reverse the pre-processing actions carried out during encoding and decompress the location map.
12) Combine all recovered RGB channels to form recovered image which will be same as the original image

IV. Experimental Results

The proposed method has been implemented using Matlab R2020a on a personal computer with an Intel Core i5 CPU, 16GB memory, with Windows 11. The performance of the proposed method was compared with Li *et al.*'s [14] and Ou *et al.*'s [15] methods using six standard 512 x 512 color images from the USC-SIPI database, including Lena, Airplane, Baboon, Lake, Peppers, and House. In addition, 24 color images of sizes 768 x 512 or 512 x 768 pixels from the Kodak image database were used to further evaluate the effectiveness of the proposed method. The performance of the proposed method was evaluated by comparing it with the performances of Li et al.'s [14] and Ou *et al.*'s [15] methods. The results show that the proposed method outperforms Li *et al.*'s [14] and Ou *et al.*'s [15] methods for Lena, Baboon, Peppers, and Airplane images, and had better performance than Ou *et al.*'s [15] method in Lake and House images, as demonstrated in the performance graphs (shown in fig.-4).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:57 UTC from IEEE Xplore. Restrictions apply.

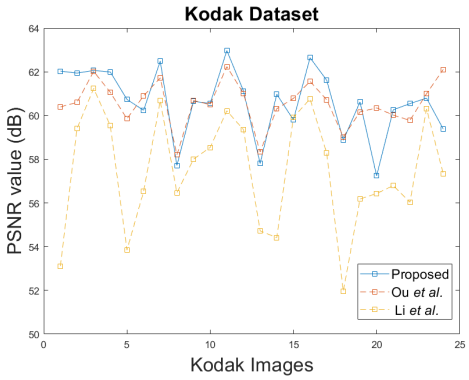


Fig. 5: Performance on Kodak Image Dataset

To further validate the performance of the proposed method, experimental results on the Kodak images are shown in fig.-5. For the comparison, the embedding capacity was set to 50,000 bits for the Kodak Graph. The results show that the proposed method outperforms Li *et al*. [14] and Ou *et al.*'s [15] methods on majority of the test images (i.e., 13 images out of 24 total images) while its performance is comparable to Ou *et al.*'s [15] method for other 3 images. However, the proposed method lacks on the remaining images with respect to Ou *et al.*'s [15] method. This is because of the large number of pixels with values of 0 or 255 which on the other hand means relatively larger size of auxiliary information and location maps.

V. Conclusion

The study introduces a PEE-based RDH method for color images. The method uses edge data from the three RGB channels, to apply an appropriate prediction approach at the current channel. Further, the use of skewed histograms for color image data embedding is also introduced with two extreme predictions. Thus, the proposed method outperformed conventional RDH works for color images that separately embed data in each channel and employ rhombus predictor for non-edge pixels, according to the presented experimental results. In future work, more focus will be on to further improve the security for providing more effective RDH method.

References

[1] R. Kumar and S. Chand, "A novel high capacity reversible data hiding scheme based on pixel intensity segmentation," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 979–996, 2017.

[2] R. Kumar, S. Chand, and S. Singh, "A reversible high capacity data hiding scheme using combinatorial strategy," *International Journal of Multimedia Intelligence and Security*, vol. 3, no. 2, pp. 146–161, 2018.

[3] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE transactions on image processing*, vol. 13, no. 8, pp. 1147–1156, 2004.

[4] S. Kim, X. Qu, V. Sachnev, and H. J. Kim, "Skewed histogram shifting for reversible data hiding using a pair of extreme predictions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 11, pp. 3236–3246, 2018.

[5] R. Kumar and K.-H. Jung, "Robust reversible data hiding scheme based on two-layer embedding strategy," *Information Sciences*, vol. 512, pp. 96–107, 2020.

[6] R. Kumar, D.-S. Kim, S.-H. Lim, and K.-H. Jung, "High-fidelity reversible data hiding using block extension strategy," in *2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*. IEEE, 2019, pp. 1–4.

[7] N. Kumar, R. Kumar, and R. Caldelli, "Local moment driven pvo based reversible data hiding," *IEEE Signal Processing Letters*, vol. 28, pp. 1335–1339, 2021.

[8] N. Kumar, R. Kumar, A. Malik, S. Singh, and K.-H. Jung, "Reversible data hiding with high visual quality using pairwise pvo and pee," *Multimedia Tools and Applications*, pp. 1–26, 2023.

[9] L. Dong, J. Zhou, Y. Y. Tang, and X. Liu, "Estimation of capacity parameters for dynamic histogram shifting (dhs)-based reversible image watermarking," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2014, pp. 1–6.

[10] X. Qu and H. J. Kim, "Pixel-based pixel value ordering predictor for high-fidelity reversible data hiding," *Signal Processing*, vol. 111, pp. 249–260, 2015.

[11] X. Li, W. Zhang, X. Gui, and B. Yang, "Efficient reversible data hiding based on multiple histograms modification," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 9, pp. 2016–2027, 2015.

[12] H.-Z. Wu, Y.-Q. Shi, H.-X. Wang, and L.-N. Zhou, "Separable reversible data hiding for encrypted palette images with color partitioning and flipping verification," *IEEE transactions on circuits and systems for video technology*, vol. 27, no. 8, pp. 1620–1631, 2016.

[13] D. Hou, W. Zhang, K. Chen, S.-J. Lin, and N. Yu, "Reversible data hiding in color image with grayscale invariance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 363–374, 2018.

[14] J. Li, X. Li, and B. Yang, "Reversible data hiding scheme for color image based on prediction-error expansion and cross-channel correlation," *Signal Processing*, vol. 93, no. 9, pp. 2748–2758, 2013.

[15] B. Ou, X. Li, Y. Zhao, and R. Ni, "Efficient color image reversible data hiding based on channel-dependent payload partition and adaptive embedding," *Signal Processing*, vol. 108, pp. 642–657, 2015.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:44:57 UTC from IEEE Xplore. Restrictions apply.

Original Research

Role of Oxidative Stress and Genetic Polymorphism of Matrix Metalloproteinase-2 and Tissue Inhibitor of Metalloproteinase-2 in COPD

Poonam Sangwan[1] Huma Quasimi[2] Jai Gopal Sharma[1] Mohammad Anwar Habib[3] Mohammad Iqbal Alam[2]

[1] Department of Biotechnology, Delhi Technological University, Delhi, India

[2] Department of Physiology, Hamdard Institute of Medical Sciences and Research, Jamia Hamdard, Delhi, India

[3] Department of Medicine, Hamdard Institute of Medical Sciences and Research, Jamia Hamdard, Delhi, India

Address Correspondence to:

Prof. Mohammad Iqbal Alam
Department of Physiology
Hamdard Institute of Medical Sciences and Research
Jamia Hamdard, New Delhi, India
Phone: 9-1-9953971178
Email: iqbalasc@yahoo.com

Running Head:

Abbreviations:
COPD - chronic obstructive pulmonary disease
MMP - Matrix metalloproteinase
ECM - Extracellular matrix
IL - Interleukin
TIMP - Tissue inhibitor of metalloproteinase
NO - Nitric oxide
MDA – Malondialdehyde

Keywords: COPD; MMP2; TIMP2; SNP; oxidative stress

Funding Support:

Date of Acceptance: July 29, 2023 | ***Publication Online Date***: August 3, 2023

Citation: Sangwan P, Quasimi H, Sharma JG, Habib MA, Alam MI. Role of oxidative stress and genetic polymorphism of matrix metalloproteinase-2 and tissue inhibitor of metalloproteinase-2 in COPD. *Chronic Obstr Pulm Dis.* 2023; Published online August 3, 2023. https://doi.org/10.15326/jcopdf.2022.0370

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

ABSTRACT

Chronic obstructive pulmonary disease (COPD), a complaint described by progressive and inadequately reversible limitation in lungs with systemic inflammation, is largely current in India. There's no remedy available so far it is, thus, imperative to understand the underpinning pathogenesis of the complainant. A set of proteases known as Matrix metalloproteinase (MMPs) are especially involved in the process of alveolar destruction and mucus hypersecretion. There are responsible factors in an inheritable position to control COPD like MMPs and TIMPs (Tissue Inhibitor of Metalloproteinases). MMPs degrade extracellular matrix and lead to the pathogenesis of COPD [1]. TIMPs proteins that help to inhibit the Matrix metalloproteinases. [2]. This review summarizes the implicit part of crucial MMP-2 and TIMP-2 in COPD disease. Though the concept seems promising, limited knowledge about the exact functions of a particular MMP in COPD and the complications of MMP in substrate affinity makes this a grueling task. MMP2 and TIMP2 both are directly or indirectly regulated by oxidative stress and epigenetic mechanism which regulates their expressions. COPD is a seditious response to factors like dust, smoke, etc., and triggers extra-pulmonary goods which cause inflammation. [3]. This review explains the relationship between MMP2 and TIMP2 in COPD patients with oxidative stress, its impact on COPD pathogenesis, and gene expression of TIMP2 and MMP2 with their downstream effects. This also gives some insights into therapeutic interventions for targeting these enzymes. MMP2 and TIMP2 both play a role in the development of COPD and they need to be studied with the utmost focus.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

1. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a major healthcare burden worldwide, although it manifests at an age generally after 40+ as part of multimorbidity, but there are evidences that there are events at an early stage of life that contribute to impaired lung functions in adults [4]. Lung tissue in COPD cases is characterized by variable and localized nonspecific changes such as habitual inflammation and tissue revamp with seditious cell infiltration, granulation tissue, thickened basement membrane and vascular wall, bronchial gland, and smooth muscle hyperplasia and hypertrophy, as well as fibrosis and epithelial metaplasia. Control lung tissue is characterized by overall low figures of cells immunoreactive for MMP-2 and TIMP-2 at different localizations. This indicates the stable, nonstop, and adaptive "birth" position of these tissue factors at a relative norm [5]. More pronounced findings of these factors in the bronchial and alveolar epithelium, and among alveolar macrophages are associated with the conservation of original lung homeostasis, including impunity and tissue reactivity; still, in hyaline cartilage, Malleability and the implicit compensatory response with the tissue increased. Significantly advanced figures of MMP-2 and TIMP-2 immunoreactive cells in COPD affected lung tissue compared to the control group indicate the part of these factors in tissue remodeling procedures. Small ascendance of TIMP-2 may limit tissue destruction processes [6].

COPD becomes worse when the walls of alveolar tissues start to damage because of its progression into Emphysema and Bronchitis and the capacity diminishes for the production of ECM [7-8]. MMP2 and TIMP2, both are very important for this progression because TIMP2 controls MMP2.

Although COPD is more common in men than women in developed countries, COPD is very prevalent in women because of high smoking rate and indoor air pollution due to biomass fuel in

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

developing countries [9]. High prevalence rate of COPD leads to high mortality and morbidity all over the world includes especially two diseases ischemic heart disease and cerebrovascular disease. However, considering that a third of COPD patients die of ischemic heart disease and a third of patients die of cardio-vascular disease with airflow limitation [10-12]. COPD disease progress includes loss of elasticity of alveoli and parenchymatous tissues leading to a decrease in Forced expiratory volume in 1 sec (FEV1) and subsequently leading to hyperinflation [13]. In addition to external factors, Genetic factors are also responsible for COPD like chronic bronchitis, loss of lung function, lung development, etc. [14, 15]. Since COPD is an inflammatory disease, it also damages DNA by Oxidative stress which leads to disturb equilibrium between tissue repair and cell proliferation [16]. Fig (1) shows that smoking leads to oxidative stress which damages DNA in COPD patients.

Smoking cessation is one kind of treatment that can be done to lower decline in lung function as well as exacerbations [17]. Patients who are addicted to tobacco, it is very hard for them to quit tobacco [18]. Arbitrations like intensive counselling, nicotine replacement therapy, and treatment with Varenicline helps to quit smoking and also much cost effective [19].

There are various genes like MMPs and TIMPs, which affects lung function and TIMPs inhibits MMPs. MMPs (Matrix Metalloproteinases) are a family of Calcium and Zinc dependent proteolytic enzymes degrades extracellular matrix which are derived from macrophages and neutrophils. Normally MMPs helps in remodeling of extracellular matrix, facilitates cell migration and are involved in immune responses by cleaving inactive forms of cytokines and chemokine's [20]. In pathologic conditions, a change in MMP expression and activity occurs which leads to destruction of extracellular matrix or tissue destruction and lung inflammation [21, 22].

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

TIMPs are Tissue Inhibitor of metalloproteinases which inhibits the activity or expression of MMPs. Basically, there are four TIMPs (TIMP1, TIMP2, TIMP3 and TIMP4) that inhibits active forms of MMPs. Although all are inhibitors of MMPs but they have different affinity for different MMPs as like TIMP2 is a potent inhibitor of MMP2 gene [23]. Distinct findings of tissue factors in bronchial epithelium, connective tissue, and blood vessels prove the involvement and significance of these structures in the morpho-pathogenesis of COPD. In general, COPD morpho-pathogenesis is characterized by increased MMP-2, TIMP-2 still, reduced Hsp-70 and hBD-4 expression that indicates high exertion of patient complex cytokine network, tissue remodeling, and overall tissue antimicrobial protection [24]. In COPD, bronchoscopically determined hypertrophy and habitual bronchitis are associated with the presence of MMP-2 in COPD-affected lung. Epithelial metaplasia and fibrosis in the histological sections are associated with worse functional parameters MMP-2. Also, the finding of granulation tissue is associated with a longer history of smoking and more pronounced tissue damage. Aging changes in the lungs at relative health are individual, indeed different for Individualities in colorful age groups, also, are substantially characterized by a lack of inflammation and variability, indeed a drop, in the number of vulnerable cells. In turn, in the case of COPD, aging is associated with worsening of the course of the complaint, as well as continuity of seditious cytokines and altered revamp with else typical COPD events. Fig 1 showing how free radicals leads to oxidative stress and DNA damage.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

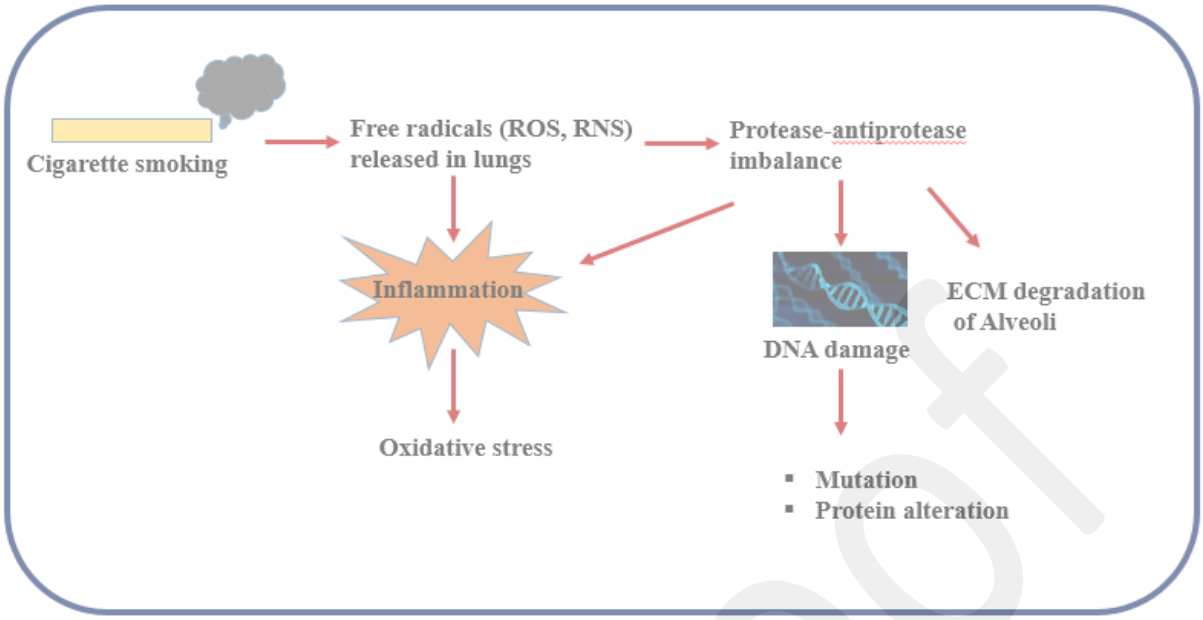


Fig 1: Cigarette smoking releases ROS and RNS in the lungs which creates proteases-antiproteases imbalance and leads to oxidative stress and DNA damage in COPD;ECM- Extracellular matrix, ROS- Reactive Oxygen species, and RNS- Reactive Nitrogen species

2. OXIDATIVE STRESS AND COPD

Oxidative stress is a potentially harmful imbalance in the oxidant–antioxidant balance in favor of the former [25]. This imbalance is because of the reactive molecules which include free radicals and non-radical reactive species. Generally, ROS includes superoxide anion and hydroxyl radical, they have unpaired electrons and their unstable nature triggers them to transfer electrons to other molecules via oxidation that leads to damage or inactivation. Lungs are vulnerable to damage by oxidative stress because of high concentrations of oxygen, increased blood supply, and exposure

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

to environmental toxins and pathogens. Cigarette and biomass smoke adds to this burden. A single puff of cigarette smoke contains 1 x 10^{15} oxidants molecules [26].

An important link between oxidative stress and the pathogenesis of COPD would come from the demonstration of the response of ROS with target lung motes and the presence of these oxidatively modified motes in increased quantities in the lungs of smokers, particularly those who develop COPD. ROS similar to O2 and OH, generated and released by activated vulnerable and seditious cells that are largely reactive and when generated close to cell membranes oxidize membrane phospholipids (Lipid peroxidation), a process that may continue as a chain response [27].

ROS directly targets proteins, DNA, and lipids which causes lung injury or generates several cellular responses via secondary metabolites or cytokines, or inflammatory molecules [28]. Fig 2 shows the different effects caused by oxidative stress.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

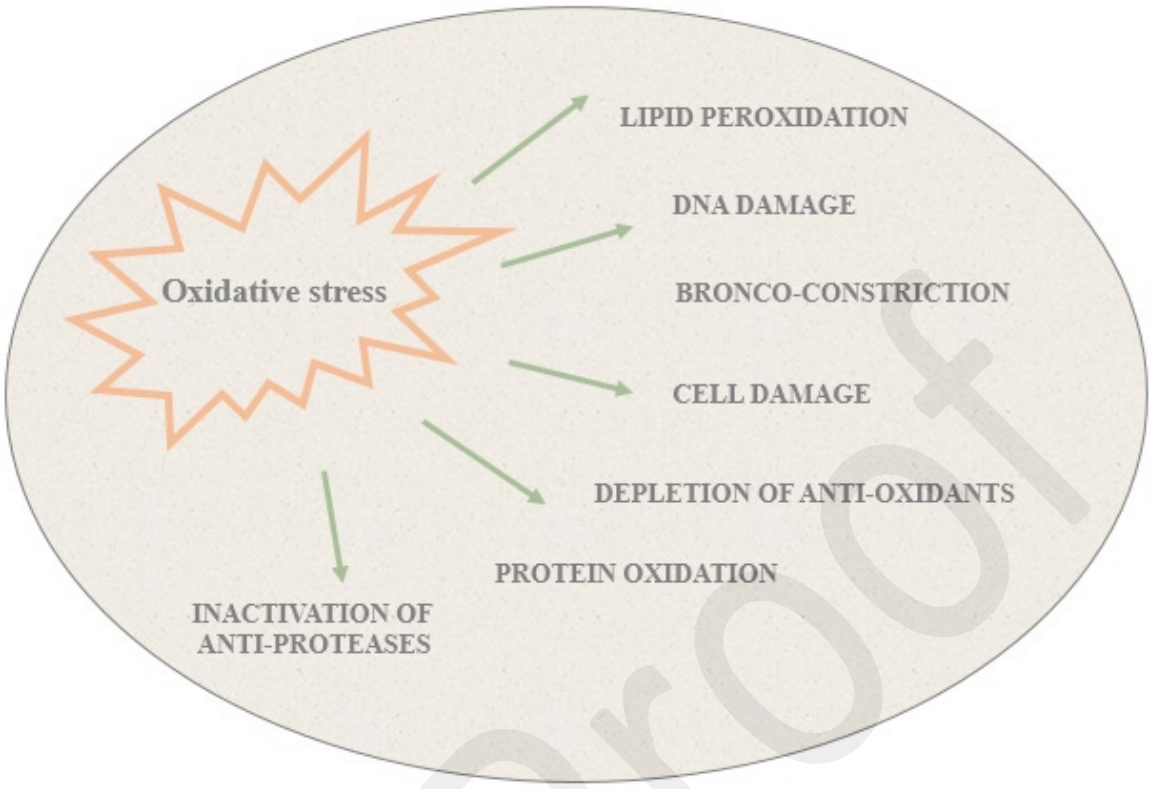


Fig 2: Different effects caused by Oxidative stress in COPD patient

Numerous effects of ROS in airways may be intermediated by the secondary release of seditious lipid intercessors similar to 4 hydroxy-2-nominal 4-(HNE), a footmark of oxidative stress/ lipid peroxidation. 4-HNE, a largely reactive diffusible end product of lipid peroxidation, is known toregulate various cellular events, similar to proliferation, apoptosis, and activation of MAPK signaling pathways [29]. 4-HNE has a high affinity towards cysteine, histidine, and lysine remainders. It forms adducts with proteins, altering their function. It also acts as a chemo-attractant for neutrophils in vitro and in vivo. Recent data indicate increased 4-HNE-modified protein situations in the airway and alveolar epithelial cells, endothelial cells, and neutrophils in subjects with airway inhibition, compared with subjects without airway inhibition [30]. Increased situations

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

of 4-HNE adducts in the lungs of cases with COPD, later showed that cigarette smoking enhanced the situations of 4-HNE adducts in bronchiolar epithelial and alveolar type II cells in mice [31].

2.1 MDA

Malondialdehyde (MDA) is a reactive organic molecule found among eukaryotes and is made up of three carbon molecules with two aldehyde groups at the C- 1 and C- 3 positions. MDA formation is induced non-enzymatically by ROS and enzymatically by lipoxygenase [32]. However, due to their instability and reactivity in both instances, the measurement of primary lipid hydroperoxide products involves a laborious task. And that is why, the quantification of lipid peroxidation is done by estimating the concentration of secondary oxidation products derived from the primary products, which are mostly aldehydes like MDA [33]. Fig 3 shows the formation of MDA from polyunsaturated fatty acids.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

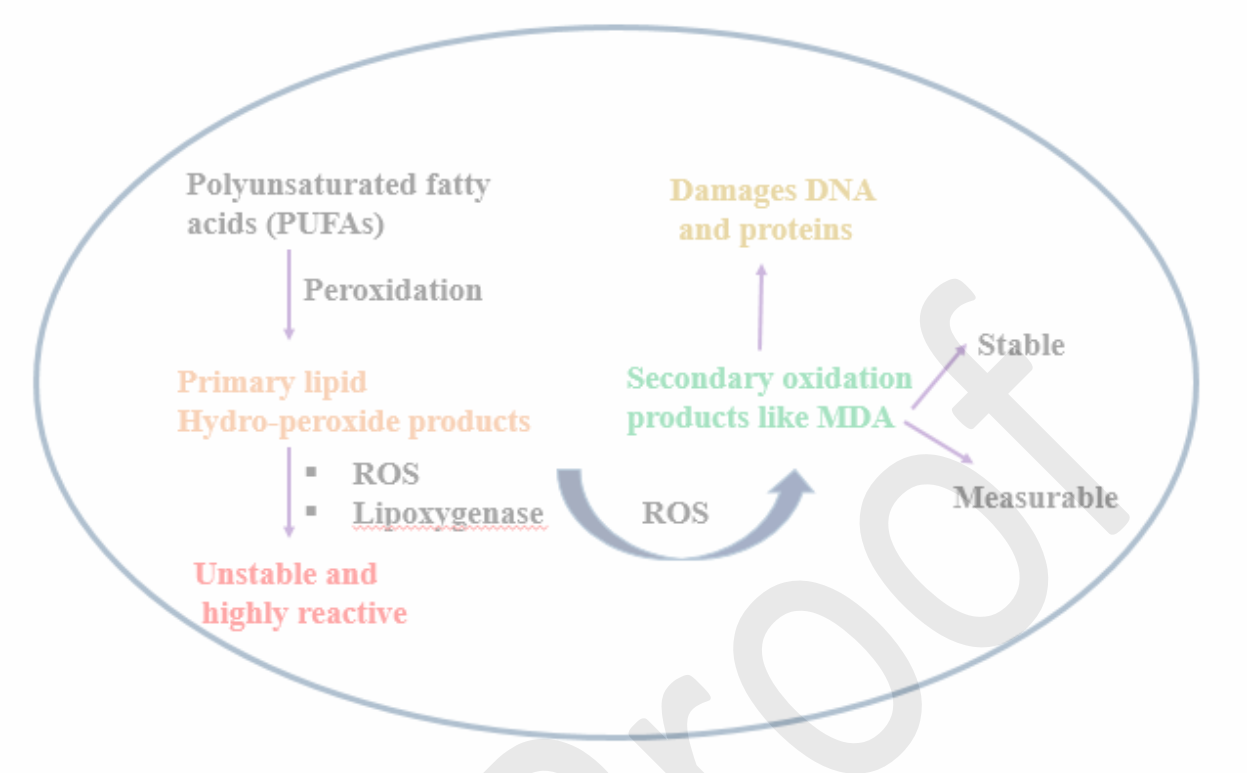


Fig-3: Lipid peroxidation from polyunsaturated fatty acids because of the high level of free radicals released in the body

MDA is widely used as a biomarker for lipid peroxidation in oxidative stress in COPD patients, in this a reaction of MDA with Thiobarbituric acid (TBA) makes MDA-TBA conjugate that absorbs light at 532 nm and gives red-pink color [34-36]. TBARS assay is considered as a good indicator of oxidative stress in biological samples [37]. Under acidic conditions (pH = 4) at 95°C, MDA bis-dimethyl acetal yields MDA. MDA and other chemicals (similar to MDA) reacts with two molecules of TBA to generate TBARS (Thio-barbituric acid reactive substances) which gives red-pink color at 532 nm [38, 39]. Generally, MDA level is elevated in COPD patients due to lipid peroxidation that further damages proteins, DNA, etc.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

2.2 GLUTATHIONE

Glutathione (GSH) is an important enzyme found in the lungs that protects our body. It is an antioxidant system. But its level is gradually reduce with continuous exposure to smoke in smokers. A high level of GSH is found in the epithelial lining fluid of cigarette smokers [40-41]. Plasma GSH level is determined using plasma, precipitated by Sulfo-salicylic acid in a 1:1 ratio. Then samples were kept at 4°C for one hour, then centrifuged at 3500 rpm for 15 min and after which the supernatant is collected. The mixture that is obtained contains 50 µl of supernatant, 50 µl of PBS (0.1 M, pH 7.4), and 10 µl of 0.4 % DTNB (5, 5′-dithiobis-(2-nitrobenzoic acid**)**. Then the chemical reaction gives yellow color which is measured at 412 nm and expressed in nm/ml of plasma [42].

2.3 SOD

SOD is an antioxidant enzyme that catalyzes the dismutation of superoxide anion to hydrogen peroxide, which is further detoxified into oxygen and water by the catalase enzyme in our body [43]. In COPD patients, the level of SOD is markedly decreased. The superoxide dismutase assay developed by Marklund and Marklund spectrophotometrically determines the SOD level in samples by checking the percent inhibition of auto-oxidation of Pyrogallol at 420 nm. SOD inhibits the auto-oxidation of Pyrogallol at an alkaline pH, with the increase in pH percentage inhibition also increases to a certain extent and SOD activity is thereby determined which is expressed in units per milliliter. Pyrogallol auto-oxidizes rapidly especially in alkaline solution. The reaction is inhibited by SOD to 99% at pH 7.9, indicating the total dependence on the participation of superoxide anion radical in this reaction. At pH 9.1 the reaction is still inhibited to over 90% by

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

SOD. But at such higher alkalinity superoxide anion radical independent mechanism rapidly becomes dominant [44]. A unit of enzyme is defined as the amount of enzyme which inhibits the reaction by 50% [45].

2.4 Nitric Oxide

Nitric Oxide (NO) is released endogenously in the lungs and acts as a dilator of bronchial and vascular smooth muscles, a neurotransmitter, and an immune response mediator [46]. NO is a marker of airway inflammation, and its level is increased in patients with COPD. Cigarette smoke increases the formation of RNS (Reactive Nitrogen species) and leads to nitration and oxidation of plasma proteins [47]. Nitric oxide is generated from L-arginine and the reaction is catalyzed with the nitric oxide synthase (NOS) enzyme. NOS is normally found in endothelium and neurons, while inducible NOS (iNOS) is primarily occurs in macrophages. NOS catalyzes the conversion of L-arginine into L-citrulline and nitric oxide.

Corradi and colleagues reported that exhaled NO is increased in stable COPD and correlated positively with lung function as assessed by forced expiratory volume in 1 sec (FEV1) [48]. Rutgers and team found that NO metabolism was not affected in stable COPD patients [49]. Another study suggested that there was no difference between COPD and healthy controls and an inverse correlation was found between FEV1 and NO levels in COPD [50]. NO was assayed using Griess reagent method in plasma samples colorimetrically. Griess reagent reacts with nitrite to form pink to a dark pink color after incubation. The absorbance of samples is determined at 540 nm and NO level is expressed in picomole/milligram of protein using $NaNO_2$ as the standard [51]. Therefore, substantiation is accumulating that cigarette bank triggers oxidative stress, leading to inflammation in the lungs of smokers and cases with COPD. Fig 4 showing that cigarette smoke

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

responsible for releasing free radicals and leads to inflammation in lungs and causes imbalance of MDA, NO, SOD and GSH etc in human body and it ultimately develops in COPD.

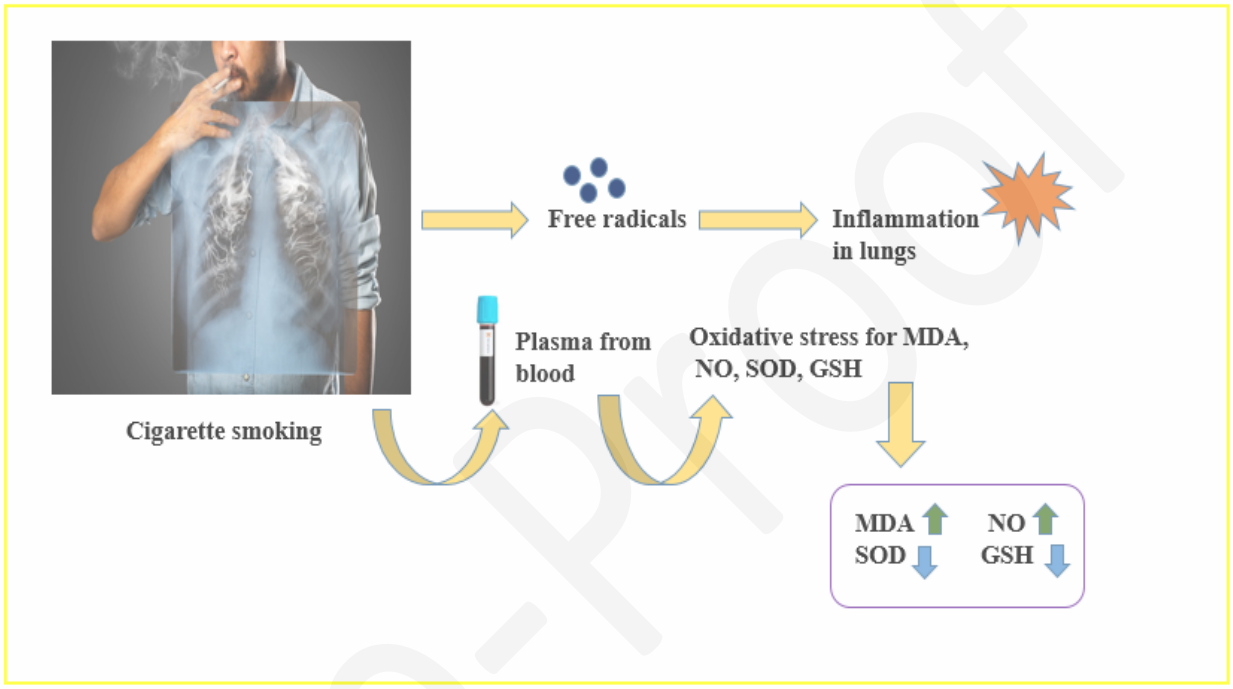


Fig 4: Oxidative stress leads to inflammation and changes the levels of MDA, SOD, GSH and Nitric oxide in human plasma

2.5 Impact of oxidative stress on COPD pathogenesis:

Oxidative stress contributes to the pathogenesis of COPD by inactivation of antiproteases, elevation of bronchial inflammation by redox; activation of neutrophils, macrophages, and fibroblasts; abnormal airway T cell population, etc. [52, 53]. Free radicals like superoxide ($^{\circ}O_2$),

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

sulfite ions, nitrogen monoxide or nitrogen dioxide radicals, hydroxyl radicals (°OH), etc are responsible for many changes in lung airways or affect different mechanisms which lead to the development of COPD. ROS may remodel ECM and blood vessels, regulates or stimulates the secretion of mucus, inactivate antiproteases, activation of transcription factors in thelungs.[54-57]. Cigarette smoking is also a crucial factor in COPD pathogenesis as the smoke contains high concentration of free radicals that accumulates neutrophils and macrophages into the lungs further leading to COPD exacerbation [58, 59]. Free radicals released from cigarette smoke cause pulmonary dysfunction [60]. Smoking also enhances 4 hydroxy-2-nonenal (4-HNE), a highly reactive diffusible end product of lipid peroxidation, it forms adducts with proteins and alters their functions in COPD [61]. Reactive nitrogen species (RNS) cause an abnormal inflammatory response in COPD by increasing nitrated proteins [62]. The antioxidant capacity of COPD also lowers because of smoking and exacerbations. Glutathione maintains the integrity of the lung airspace epithelial barrier but decreases the level of GSH in the epithelial leads to loss of barrier function and permeability [63, 64].

Many signal transduction or mechanisms are affected by oxidative stress which plays an important role in COPD pathogenesis such as the inactivation of antitrypsin α1-AT by oxid ation of methionine in its active site of the protein, lead to the low level of protein and lung airways are damaged by proteases [65].

3. MMP2 and TIMP2

MMP2 is a type of Gelatinase A, 72 KDa protein released by smooth muscle cells and epithelial cells in the lungs as well as by lymphocytes and macrophages by infiltrating the tissue [66].It helps in vascularization, endometrial menstrual breakdown, and inflammatory activities [67].

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

Extracellular matrix (ECM) provides mechanical support to cells and is a complex of different proteins like collagen, laminins, cytokines, fibronectins, and growth factors but MMPs proteolytic activity degrades ECM and also provides signals to embedded cells to react to stimuli [68]. MMP2 also cleaves non-ECM molecules like pro-TNF-α, pro-IL-1β, pro-IL8, MCP-3, etc. and can also activates various MMP Zymogens as pro-MMP-1, pro-MMP-2, pro-MMP-9 and pro-MMP-13 [69-72].

MMP2 contains polymorphic sequences in its promoter site like Sp-1 binding site. Sp-1 binding site determines the basal activity of MMP2 promoter [73]. It has been proved that the polymorphism in the promoter site of MMP2 results in change in expression of MMP2 gene [74].

TIMPs (Tissue inhibitor of Metalloproteinase) is present in four active forms (TIMP1, TIMP2, TIMP3 and TIMP4) which inhibits the activities of MMPs by binding to their active sites [75]. TIMP2 is a potent inhibitor of MMP2. It was also seen that TIMP2 is involved in pulmonary disease characterized by abnormal tissue remodeling of alveolar structure in lungs like emphysema, fibrosis, acute respiratory distress syndrome and lung cancer [76]. TIMP2 promotes cell growth and protects cell from apoptosis. TIMPs also shape and organize appropriate ECM environment and architecture [77, 78]. Fig 5 depicts MMP2-TIMP2 regulation and degradation of ECM by MMP2. .

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

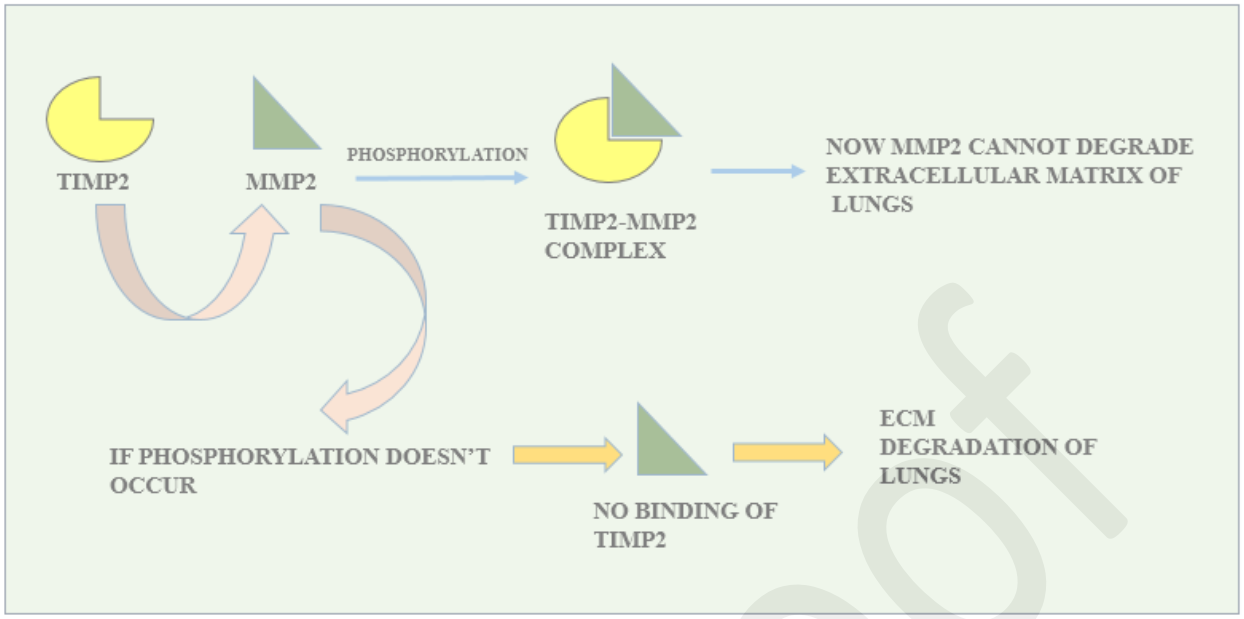


Fig-5: Regulation of MMP2 by TIMP2 as TIMP2 inhibits MMP2 by phosphorylation but without its binding, MMP2 degrading ECM of lungs;TIMP2- Tissue Inhibitor of Metalloproteinase 2, MMP2- Matrix Metalloproteinase 2

Both Bronchial epithelial cells and airway muscle cells secrete MMP2, and smoke or dust trigger the up-regulation of MMP2 protein in COPD cases [79]. TIMP2 is a potent asset of MMP2 and it binds with its active site after phosphorylation on three different sites by the c-Src kinase enzyme. After phosphorylation, TIMP2 exclusively binds with MMP2 and inhibits its exertion [80].

3.1 Importance of TIMP2 and MMP2 in COPD:

MMP2 is a Gelatinase A enzyme responsible for COPD's pathogenesis. It is released from macrophage and bronchiolar epithelial cells and requires TIMP2 for its activation [81, 22]. MMP2 causes pathological changes in the pulmonary interstitium and lung function of COPD [83] and is

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

also involved in ECM remodeling and interstitial fibrosis [84]. TIMP2 inhibition by of MMP2, is involved in abnormal ECM accumulation [85, 86]. TIMP2 binds non-covalently to MMP2 with the help of MMP14 and regulates their activities. TIMP2 binds pro-MMP2 molecule (inactive) with MMP14 molecule. The N-terminal of TIMP2 binds with the active site of MMP2 and inactivating it by forming an inhibitory complex. The binding of TIMP2 to MMP14 allows the cleavage of pro-MMP2 by a free MMP14 molecule, as an active MMP2 molecule [87-89]. Loss of TIMP2 leads to a decrease in the activation of pro-MMP2 [90]. TIMP2 is inhibited when it undergoes a mutation or when free radicals are produced. And because it is in charge of regulating MMP2 production, MMP2 will become greatly elevated, begin digesting the ECM in the lungs, and cause COPD. The remodeling of the lung matrix, which is important for the pathogenesis of COPD, is carried out by MMP2 and TIMP2. [91, 92].

4. MMP2 and TIMP2 at Global level:

MMP2 Single nucleotide polymorphisms (SNPs) have been seen among various populations with TIMP2 SNPs worldwide. Different SNPs have been studied all these years.

MMP2 -1306 at promoter site (rs243865) in central Bulgaria has been studied and that included 71 individuals.. This study showed that the genotype but not allele frequency was different between COPD patients and controls. The carriers of T alleles were at higher risk for developing COPD [93]. MMP2 rs243864, rs3918253, and rs11646643 SNPs were associated with a high risk of COPD in the Mexican population [94]. MMP2 rs2285053 and rs243865 SNPs were also found

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

to be associated with the development of COPD or lung carcinoma in Asians but not the Caucasians population [95].

TIMP2 gene polymorphism at two different positions +853 G/A and -418 G/C showed a down-regulation of TIMP2 in COPD patients than controls in the Japanese population which lead to the up-regulation of MMP2 in patients for degradation of ECM in COPD patients. G/A nucleotide substitution at +853 in exon 3 and G/C substitution at -418 in Sp1 binding site in the Japanese population, showed elevated genotypic and allelic frequency at +853 position and higher allele frequency at -418 position between COPD patients and the control group [96].

But these two SNPs differed in the Egyptian population, +853 G/A showed significant genotype difference in controls and COPD patients but there was no difference between controls and COPD patients in -418 G/C SNP [97]. Table 1 summarizes different SNPs of TIMP2 and MMP2 genes in different regions worldwide studied in COPD patients. MMP2 and TIMP2 were also found to be associated with chronic bronchitis in Tatar children [98].

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Gene	SNPs	Location	Substitution	References
TIMP2	+853	Exon 3	G/A	[97]
	-418	Sp1 binding site	G/C	[97]
MMP2	rs243865	-1306 Promoter	C/T	[74]
	rs243839	Intron	G/A	[94]
	rs243835	Intron	C/T	[94]
	rs243864	-790 Promoter	T/G	[94]
	rs11646643	Intron	A/G	[94]
	rs2285053	735 Promoter	C/T	[95, 98]

Table 1: Different SNPs of TIMP2 and MMP2

COPD, a common respiratory complaint, is one of the leading causes of morbidity and mortality

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

among smokers and some non-smokers [99]. India contributes a significant and raising chance of COPD mortality which is evaluated to exist among the loftiest in the people; crude assessments indicate there are 30 million COPD cases in India [100].

MMP-2, an important modulator of cell proliferation, is buried by both, bronchial epithelial cells and airway smooth muscle cells. Increased vulnerable reactivity for MMP-2 in COPD lungs , substantially in alveolar macrophages and airway epithelial cells was observed [101]. Increased MMP-2 exertion in mice exposed to long-term cigarette banks was also reported. MMP-2 mRNA up-regulation and increased protein in lavage and lung towel was also linked to emphysema convinced by the wood bank [102]. In discrepancy, a study using ray prisoner micro-dissected samples of mortal lung parenchyma has shown that MMP-2 gene expression situations drop with the inflexibility of the patient [103].

However , TIMPs haven't been considered as remedial targets for COPD due to their variation in affinity for different proteases. Though the part of TIMPs in cranking pro-MMPs is well accepted, several studies live that reported the MMP-limited function of TIMPs, similar to inhibition of the mitogenic response of mortal endothelial cells to growth factors and creation of apoptosis [104-106].

5. Impact of oxidative stress on the gene expression or the genetic mutations of TIMP2 and MMP2:

Oxidative stress because of cigarette smoking or biomass fuel affects the genetic makeup of the MMP2 and TIMP2 gene which alters their protein product and is responsible for COPD pathogenesis. According to a study conducted on the Japanese population, the polymorphism of +853 G/A and -418 G/C SNPs in the TIMP2 gene was higher in smokers. Smoking destabilizes

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

the TIMP2 gene and causes its low production [107, 108]. The same polymorphism of +853 SNP was seen in an Egyptian population associated with COPD in smokers [109].

MMP2 is mainly expressed in mesenchymal cells (mainly fibroblasts) during development and tissue regeneration. Due to airway remodeling, there are structural changes with an inflammatory response. There are many promoter polymorphisms in the MMP2 gene that alter their expressions such as -1306 C>T SNP. This gene expression is altered because smoking and oxidative stress lead to inflammation and ECM degradation which affects this gene at the genetic level [110]. Smoking causes the release of various oxidative factors or imbalances oxidant or antioxidant, or proteinase or anti-proteinase enzymes which alters the genetic makeup of the MMP2 or TIMP2 gene. This causes a change in their downstream products at the transcriptional or translational level and changing their functions and leading to the development of COPD.

Most of the MMPs share common cis-elements in their promoters, which are regulated by various cytokines like (IL-1b, TNF-α, TGF-β), growth factors (EGF, VEGF, bFGF), oxidative stress factors, xenobiotics [111, 112]. This stimulus induces signal transduction pathways leading to the activation of transcriptional factors such as NF-κb, STATs, ETS factors, AP-1, etc [111, 113]. In this respect, the naturally occurring DNA sequence changes in the regulatory regions especially promoters of the genes encoding MMPs affect the transcriptional activity and influence the balance between MMPs and TIMPs thus causing development and progression of COPD [114].

6. The impact of oxidants on MMP2 and TIMP2:

Oxidants are free radicals (reactive forms of oxygen and nitrogen species) that are highly unstable, reactive, and tend to make reactions with different molecules, for example, hydroxyl radical OH°, superoxide, hydrogen peroxide, nitric oxide, etc. Nitric oxide (°NO_2), a reactive free radical cause

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

epithelial damage and induce inflammation in the airways which ultimately affects the MMP2 secretion that is upregulated due to the inflammation and downregulates TIMP2 in the airways. Reactive Oxygen Species (ROS) react with thiol groups that preserve MMP latency (115) thus activating MMPs. ROS enhances the expression of MMP2 via xanthine oxidase (116, 117) and via mechanical stretch-induced vascular NADPH oxidase (118). $^{-}$OONO (produced by the reaction of nitric oxide and superoxide) produces posttranslational modifications of the MMP2 structure, thus resulting in an active full-length MMP2 enzyme (119, 120). Radiation produces ROS including $O_2^{\circ-}$ and H_2O_2 activates MMP2 (121). H_2O_2 also activates the Ca^{2+} ATPase activity of MMP2 (as MMP2 is Ca dependent enzyme) in smooth muscle membranes with Vitamin E preventing the increase in activity of MMP2 and Ca^{2+} ATPase activity (122).

ROS generated by xanthine and xanthine oxidase system (XOD) reduces TIMP2 mRNA level, inhibited by catalase scavenging. ROS especially affects the stability of mRNA of MMP2 and TIMP2 (123). There are various other ROS which affects TIMP2 at transcriptional or translational level.

7. Impact and downstream effects of the TIMP2 and MMP2 gene polymorphism:

The expression levels of MMPs are affected by polymorphisms in promoter regions of genes encoding these enzymes. MMP2 is a type IV collagen, and can also degrade elastin molecules. There are three polymorphisms in the promoter region of this gene. Two of them are at -1306 and -735 positions and the third SNP is at -1575 position. The previous two SNPs cause disruption of the Sp1 regulatory element and strikingly decreased promoter activity in macrophages and

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

epithelial cells. The third SNP at -1575 is located immediately 5' to an estrogen receptor (ER) binding site and it functions as an enhancer and leads to increased transcriptional activity and also showed linkage disequilibrium and demonstrated an additive reduction in estrogen-dependent reporter activity [124]. MMP2 -1306TT genotype showed to have excess FEV1 decline and a lower mean of FEV1% compared to the wild-type carriers [125]. The extent of inflammation correlated with the reduction in FEV1 and FEV1/FVC ratio, and probably with the accelerated decline in FEV1 characteristics of COPD [126]. The lower promoter activity of the T allele may lead to increased deposition of connective tissue in the lungs and the development of airflow obstruction, which may be the reason for an increased risk for the development of COPD.

The genetic polymorphism of the TIMP2 gene at +853 G/A (at exon 3 in the TIMP2 gene) and -418 G/C (in the promoter region of TIMP2) SNPs supports the possibility of the development of COPD in different populations. This +853 SNP can down-regulate the TIMP2 activity by changing the secondary structure of the mRNA which inhibits the ribosomal binding and/or decreases the mRNA stability [127 , 128]. MMP2/ TIMP2 system plays an important role in the formation of pulmonary emphysema. So, the mutations of the TIMP2 gene down-regulate its own activity which results in the progression of the breakdown of the lung matrix [129].

Another nucleotide substitution at the -418 position, located in the consensus sequence for the Sp1 binding site in the promoter region. There may be the possibility that this substitution results in the downregulation of the transcriptional activity of the TIMP2 gene and increased activity of MMP2 which results in the development of COPD [130].

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

8. Mechanism:

Cigarette smoke is responsible for the neutrophil and macrophage influx to the airways in the lungs for NETosis (Neutrophil Extracellular Trap formation) [131-133]. Neutrophil extracellular traps are neutrophils extrude webs of decondensed chromatin with histones, neutrophil elastase, EN-RAGE (Extracellular newly identified receptor for advanced glycation end products binding protein), and myeloperoxidase [134 -136]. NETs were observed in stable and severe COPD patients and correlated with disease severity as well [137 , 138]. EN-RAGE is responsible for the downstream signaling cascade for the pro-inflammatory cytokines in this disorder [139]. Lytic NETosis induces the MAPK/ERK pathway which leads to the production of NADPH and Hydrogen peroxide, it causes the production of Neutrophil Elastase which ultimately leads to the production of ROS and RNS [135]. Leukotriene B4 (LTB4) triggers influx of neutrophils in airways and cause exacerbation in COPD patients and disease progression [140]. Macrophage and neutrophil influx increase in the pulmonary sections.

Early growth response gene product 1 (EGR1) in endothelial cells, binds to the MM14 promoter and it induces its expression. In turn, MMP14 activates MMP2 and upregulates its production [141]. Also, some transcription factors like GATA1 (Globin transcription factor 1), AP1 (activator protein 1), STAT3C (Signal transducer and activator of transcription 3), and NFκB (Nuclear factor kappa B) regulate the expression of MMP2. Tumor necrosis factor alpha (TNFα), Interferon-gamma (IFN-γ), and Interleukin 1β (IL-1β) also triggers the upregulation of MMP2 in lungs [142

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

]. IL-1β also causes pulmonary inflammation by triggering metalloproteinases in epithelial cells [143]. Also, IL13 is a pro-inflammatory cytokine that causes hyper-expression of mRNA of MMP2 in the lungs [144]. Reactive oxygen species (ROS) and RNS produced by cigarette smoke, and hydrogen peroxide in mitochondria cause tissue damage in the lungs and increase MMP2 gene expression. RNS leads to inflammation by increasing MMPs expression in the lungs [145-147]. So, all these factors causes inflammation in lungs and hyperexpression of MMP2 that primarily causing degradation of ECM of the lungs. Given that, TIMP2 is an inhibitor of MMP2. The expression of MMP2 if found higher than TIMP2, suggests that TIMP2 is inhibited or down-regulated, further indicating the rationale behind the damage of lungs ECM or alveoli or hyper mucous secretion. NETosis also causes MMP2 mutation which is responsible for various changes in gene expression and responsible for causing COPD in a particular population. Environmental factors like air pollution, dust or passive smoke, biomass fuel also causes the production of ROS/RNS in our body and ultimately leads to inflammation. Macrophage releases MIF (Migration inhibitory factor) cytokine which also causes inflammation in the lungs.

Pre-Proof

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

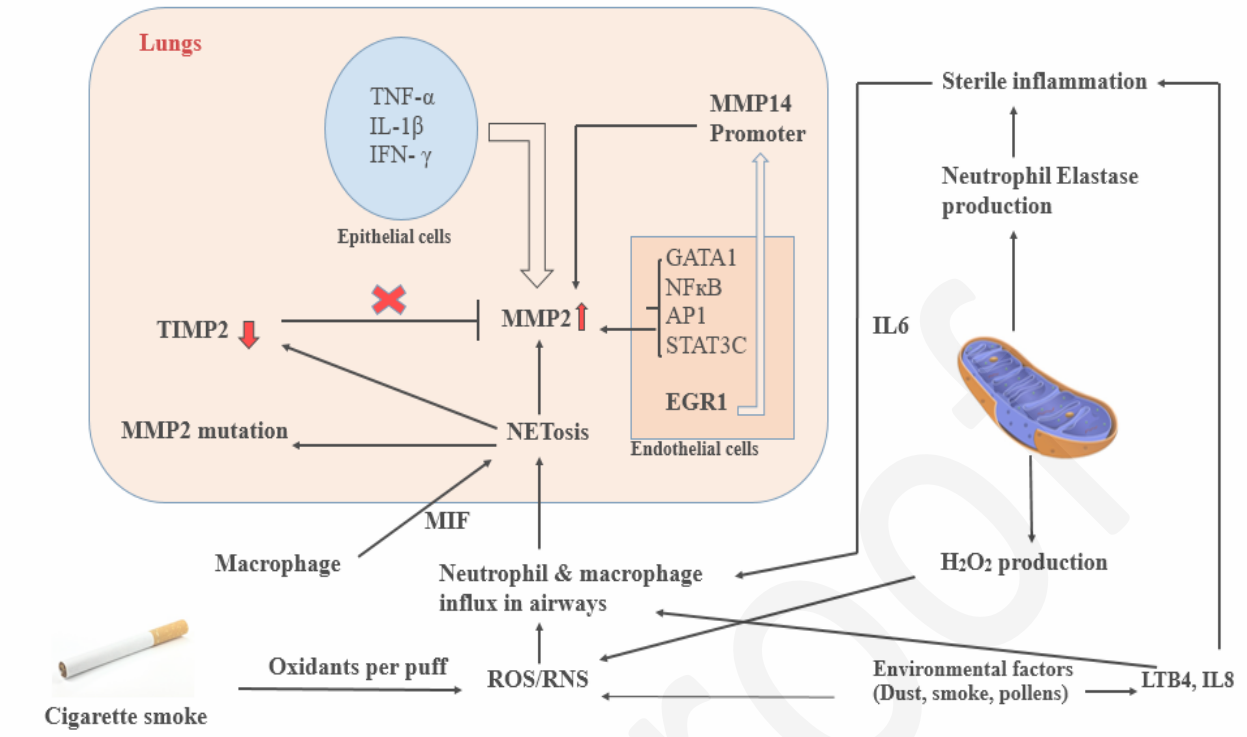


Fig-6: Mechanism showing different transcription factors, cytokines, and environmental factors causing damage in lungs or changes in the expression of MMP2. NETosis leads to mutation, inflammation in lungs which indirectly inhibits TIMP2 and upregulates MMP2.

9 Other mechanisms affecting TIMP2 and MMP2 expressions such as epigenetic changes:

Epigenetic changes are stable changes in cell function or environmental behavior that affect the way our gene works but without the change in their genetic sequence. MMP2 activity is regulated at the transcriptional level by methylation, histone modification, non-coding RNA regulation, and chromatin remodeling [148]. There are various factors like smoke, dust, pollution, biomass fuel, etc. which directly or indirectly cause epigenetic changes in various cytokines like TNF-α, IL8, HNE, NO, etc. [149 , 150] at the transcriptional or translational level by acetylation, methylation, or phosphorylation which might be responsible for the changes in MMP2 or TIMP2 expression

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

but currently there are no such reports that present any particular changes targeting these two enzymes in COPD.

10. Possible potential therapeutic intervention targeting TIMP2 or MMP2 nowadays or in the future:

Over or under-expression, Imbalance between TIMP2 & MMP2 may be corrected by usage of drugs. . Synthetic MMP inhibitors were used in preclinical and clinical studies. MMP inhibitors (Batimastat, Marimastat, Prinomastat, Neovastat, CMT-3, BMS-275291) were found efficacious against MMP-1, 2, 3, 9, 12 in experimental models of lung, breast, and colon cancer with reduced tumor growth by blocking tumor angiogenesis [151].

Simvastatin modulated the expression of MMP2 in lung cancer. It plays a key role in the prevention and treatment of lung cancer in COPD [152]. Compounds 25 and 26 can be considered good therapeutic agents for treating COPD emphysema. Inhibitors targeting specific proteases (MMPs) or protease inhibitors (TIMPs) can be used to treat COPD. These inhibitors can be used in combination with the conventional COPD drugs like ICS, LABA, Tiotropium.. MMP2 or TIMP2 can also be targeted by using specific antibodies or targeted potent inhibitors but there are no evidences that targets specifically these enzymes as yet. Currently, There are not much information present regarding targeting of MMP2 or TIMP2 enzymes in treating COPD.

11. Future prospects:

COPD and protease-antiprotease imbalance is very well known. Proteases degrades lung tissues and leads to emphysema or chronic bronchitis. TIMP2 inhibits MMP2 so that it cannot degrades extracellular matrix of lungs but when it fails because of the oxidative stress or any other genetic

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

reason, it leads to activation of MMP2. Numerous studies reported the relation of MMP2 and TIMP2 in different populations but , it has not been reported in North-Indian population till date. Do you think, Is this the same scenario in Indian population also? Many studies have been performed in the Indian population, but till date nobody correlated MMP2-TIMP2 at genetic level. So, will it give the same result? Also NET traps causing inflammation in lungs of COPD and increases exacerbation also responsible for severity of disease. So, NET traps might be a factor causing this illness. So we can also target to prevent NET traps in lungs or develop a medicine to treat these traps which is not available till date.

12 . DISCUSSION

COPD is an inflammatory disease, it has various responsible factors like dust, smoke, genetic makeup, etc. COPD is directly or indirectly affected by the imbalance created in the human body because of these inflammatory markers and leads to oxidative stress. Oxidative stress is created generally because of the consumption of cigarette by men and biomass fuel by women in villages or local areas where hukkas and chullas are used. In our body, free radicals leads to release some inflammatory molecules like IL8, IL6, TNF-alpha, etc. cytokines which ultimately causes the imbalance between oxidants and anti-oxidants which disturbs body's equilibrium and push our body system towards oxidative stress and responsible for the low level of SOD, GSH, NO and

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

high level of MDA in COPD patients but there are few reports which suggested that they have found high level of SOD and NO in COPD patients [70, 74]. Biomass fuel is also responsible for the COPD in females in villages and small towns in India. Oxidative stress is also a key factor in DNA damage. DNA damage caused by enzymes leads to mutation in DNA and protein change.

This is well known fact that matrix-metalloproteinases degrades the lung epithelial cells until or unless they are inhibited by tissue inhibitors like TIMPs. So, TIMPs and MMPs are closely related with each other. TIMPs inhibits the MMPs enzymes but in diseases like COPD, these MMPs are inhibited by TIMPs as well as the inflammation created in human physiological system, so they are hyper-regulated and starts to eat lung's walls. It degrades extracellular matrix of lungs. Various SNPs were studied in different regions like Asians, Egyptians, etc. and found different results for polymorphism of these SNPs in COPD and control groups. Different SNPs were studied of MMP2 and TIMP2 genes, where -1306 MMP2 SNP was having different genotype in Central Bulgaria population but not different allele frequency but MMP2 rs243864, rs3918253, rs11646643 SNPs and MMP2 rs2285053 and rs243865 were associated with the risk of COPD in Mexican and Asian population respectively. TIMP2 +853 and -418 SNPs studied in Japanese population, showed the low level of TIMP2 protein in COPD patients. Also there was difference in Egyptian population, where +853 SNP in TIMP2 was showing differences in COPD and Control group but not in -418 SNP. Mechanism showing different possible pathways which could be a triggering point in severity of disease or cause of disease. All these factors are very important which we can study or do some experiments for treating COPD.

So, there are various views regarding these genotype or allelic differences in COPD and control groups of different populations all over the world. Many have seen differences but few have not.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

We are going to correlate the MMP2 and TIMP2 genetic differences in COPD and control groups of the North Indian population for the very first time. Till date, to the best of our knowledge, it is not reported anywhere.

13. CONCLUSION

COPD is a respiratory disease which is caused by smoke, dust, biomass fuel, cigarette, etc. and also due to oxidative stress which is triggered by the inflammatory molecules. These inflammatory molecules lead to severity of disease in patients. Proteases-antiproteases imbalance also plays a key role in COPD development. MMP2-TIMP2 are related to each other because TIMP2 is an inhibitor of MMP2 and in case of COPD, when TIMP2 is down-regulated, MMP2 starts to digest the extracellular matrix and leads to respiratory dis-comfort in patient. Recapitulating, MMP-2 appears to play an important part in the airway ECM remodeling in COPD. It appears that numerous MMPs are over-regulated in alveolar macrophages, both in humans with emphysema and in cigarette bank-convinced mouse and guinea gormandizer models; and over-expression of some MMPs is significant. Since MMPs retain an active catalytic point and natural *in vivo* impediments, they incontinently become therapeutic targets. A wide range of MMP inhibition may be useful for minimal inhibition of ECM declination. Monoclonal antibodies against MMPs, targeting the substrate of MMPs, can alike be capitalized to clock productive MMP inhibition. Combinatorial therapy , similar to an MMP asset plus a β2 agonist or low-cure steroids could be another strategy for remedial operation of the complaint. Despite rigorous scientific trouble, the discovery of applicable MMP assets for COPD remains a challenge. It is not easy to develop MMP

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

impediments that are largely picky and specific thereby targeting only applicable MMPs. It's pivotal to understand the successional expression and individual contribution of each MMPs in the initiation and progression of COPD to allow for the development of a picky and specific target. Also, systemic toxin, lack of correlation between the exertion of the asset and MMP situations in the tube, and poor efficacy are some of the issues which need attention. Defensive impunity and towel form should also be a matter of concern while considering MMP inhibitory therapeutic strategy.

DECLARATION OF INTEREST:

We hereby declare that the manuscript has not been previously published in any language anywhere and that is not under simultaneous consideration by another journal. None of the authors have any conflict of interest or any financial ties to disclose.

AUTHORS CONTRIBUTION:

PS and HQ contributed to the study conception and design; data acquisition, analysis, and interpretation; drafting of the manuscript and final approval of the version to be published. JGS, AH and MIA contributed to the study conception and design; data acquisition, analysis, and interpretation; drafting of the manuscript; critical revision of the manuscript for important intellectual content; and final approval of the version to be published.

References:

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

1. Shapiro S D. The macrophage in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1999; 160:S29-32.
2. Murphy G. Tissue inhibitors of metalloproteinases. Genome Biol 2011; 12, 233.
3. Markewitz B A, Owens M W, Payne D K. The pathogenesis of the chronic obstructive pulmonary disease. Am J Med Sci 1999; 318:74-8.
4. Martinez FD. Early-life origins of chronic obstructive pulmonary disease. New England Journal of Medicine. 2016 Sep 1; 375(9):871-78.
5. Fischer B.M., Pavlisko E., Voynow J.A. (2011). Pathogenic triad in COPD: oxidative stress, protease-antiprotease imbalance, and inflammation. Int J Chron Obstruct Pulmon Dis, 6, 413–421. https://doi.org/10.2147/COPD.S10770
6. Srivastava P.K., Dastidar S.G., Ray A. (2007). Chronic obstructive pulmonary disease: role of matrix metalloproteases and future challenges of drug therapy. Expert Opin Investig Drugs, 16, 7, 1069–1078. https://doi.org/10.1517/13543784.16.7.1069
7. Harju T, Kinnula VL, Paakko P et al (2010) Variability in the precursor proteins of collagen I and III in different stages of COPD. Respir Res 11:165.
8. Vestbo J, Hurd SS, Rodriguez-Roisin R (2012) The 2011 revision of the global strategy for the diagnosis, management and prevention of COPD (GOLD)—why and what? Clin Respir J 6:208–214.
9. Adeloye D, Chua S, Lee C, et al. Global and regional estimates of COPD prevalence: systematic review and meta-analysis. J Glob Health 2015; 5: 020415.
10. Anthonisen NR, Skeans MA, Wise RA, Manfreda J, Kanner RE, Connett JE. The effects of a smoking cessation intervention on 14.5-year mortality: a randomized clinical trial. Ann Intern Med 2005; 142: 233–39.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

11. Calverley PM, Anderson JA, Celli B, et al. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. N Engl J Med 2007; 356: 775–89.

12. Franssen FM, Soriano JB, Roche N, et al. Lung function abnormalities in smokers with ischemic heart disease. Am J Respir Crit Care Med 2016; 194: 568–76.

13. O'Donnell DE. Hyperinflation, dyspnea, and exercise intolerance in chronic obstructive pulmonary disease. Proc Am Thorac Soc 2006; 3: 180–84.

14. Obeidat M, Hao K, Bosse Y, et al. Molecular mechanisms underlying variations in lung function: a systems genetics analysis. Lancet Respir Med 2015; 3: 782–95.

15. Wain LV, Shrine N, Miller S, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med 2015; 3: 769–81.

16. Houghton AM, Mouded M, Shapiro SD. Common origins of lung cancer and COPD. Nat Med 2008; 14: 1023–24.

17. Anthonisen NR, Skeans MA, Wise RA, Manfreda J, Kanner RE, Connett JE. The effects of a smoking cessation intervention on 14.5-year mortality: a randomized clinical trial. Ann Intern Med 2005; 142: 233–39.

18. Jimenez-Ruiz CA, Andreas S, Lewis KE, et al. Statement on smoking cessation in COPD and other pulmonary diseases and in smokers with comorbidities who find it difficult to quit. Eur Respir J 2015; 46: 61–79.

19. Hoogendoorn M, Feenstra TL, Hoogenveen RT, et. al. Long-term effectiveness and costeffectiveness of smoking cessation interventions in patients with COPD. Thorax 2010; 65: 711–18.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

20. Elkington PTF, Friedland JS. Matrix metalloproteinases in destructive pulmonary pathology. Thorax 2006; 61:259 –266.

21. Gueders MM, Foidart JM, Noel A, et al. Matrix metalloproteinases (MMPs) and tissue inhibitors of MMPs in the respiratory tract: potential implications in asthma and other lung diseases. Eur J Pharmacol 2006; 533:133–144.

22. Shapiro SD. The macrophage in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1999; 160:S29-S32.

23. Howard EW, Bullen EC, Banda MJ: Preferential inhibition of 72- and 92-kDa gelatinases by tissue inhibitor of metalloproteinases-2. J Biol Chem 1991, 266(20):13070-13075.

24. Vitenberga, Zane et al. (2017) "COPD affected lung tissue remodelling due to the local distribution of MMP-2, TIMP-2, TGF-β1 and HSP-70. Papers on Anthropology XXVI/2, 2017, pp. 157–167. DOI: 10.12697/poa.2017.26.2.16.

25. Sies H. Oxidative stress: oxidants and antioxidants. Exp Physiol. 1997 Mar; 82(2):291-5. doi: 10.1113/expphysiol.1997.sp004024. PMID: 9129943.

26. Pryor, W.A.; Stone, K. Oxidants in cigarette smoke. Radicals, hydrogen peroxide, peroxynitrate, and peroxynitrite. Ann. N. Y. Acad. Sci. 1993, 686, 12–27.

27. Ahmad A, Shameem M, Husain Q. Altered oxidant-antioxidant levels in the disease prognosis of chronic obstructive pulmonary disease. Int J Tuberc Lung Dis. 2013 Aug; 17(8):1104-9. doi: 10.5588/ijtld.12.0512. PMID: 23827037.

28. Rahman I, MacNee W. Role of transcription factors in inflammatory lung diseases. Thorax. 1998 Jul; 53(7):601-12. doi: 10.1136/thx.53.7.601. PMID: 9797762; PMCID: PMC1745272.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

29. Uchida K, Shiraishi M, Naito Y,et. al. Activation of stress signaling pathways by the end product of lipid peroxidation. 4-hydroxy-2-nonenal is a potential inducer of intracellular peroxide production. J Biol Chem. 1999 Jan 22; 274(4):2234-42. doi: 10.1074/jbc.274.4.2234. PMID: 9890986.

30. Rahman I, van Schadewijk AA, Crowther AJ,et. al. 4-Hydroxy-2-nonenal, a specific lipid peroxidation product, is elevated in lungs of patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2002 Aug 15; 166(4):490-5. doi: 10.1164/rccm.2110101. PMID: 12186826.

31. Aoshiba K, Koinuma M, Yokohori N, Nagai A. Immunohistochemical evaluation of oxidative stress in murine lungs after cigarette smoke exposure. Inhal Toxicol. 2003 Sep; 15(10):1029-38. doi: 10.1080/08958370390226431. PMID: 12928978.

32. Farmer EE, Mueller MJ. ROS-mediated lipid peroxidation and RES-activated signaling. Annu Rev Plant Biol. 2013; 64:429-50. doi: 10.1146/annurev-arplant-050312-120132. Epub 2013 Feb 28. PMID: 23451784.

33. Davey MW, Stals E, Panis B, et. al. High-throughput determination of malondialdehyde in plant tissues. Anal Biochem. 2005 Dec 15; 347(2):201-7. doi: 10.1016/j.ab.2005.09.041. Epub 2005 Oct 21. PMID: 16289006.

34. Ohkawa H, Ohishi N, Yagi K. Reaction of linoleic acid hydroperoxide with thiobarbituric acid. J Lipid Res. 1978 Nov; 19(8):1053-7. PMID: 103988.

35. Khoubnasabjafari M, Soleymani J, Jouyban A. Avoid using spectrophotometric determination of malondialdehyde as a biomarker of oxidative stress. Biomark Med. 2018 Jun;12(6):551-554. doi: 10.2217/bmm-2017-0437. Epub 2018 Jun 6. PMID: 29873547.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

36. Morales M, Munné-Bosch S. Malondialdehyde: Facts and Artifacts. Plant Physiol. 2019 Jul; 180(3):1246-1250. doi: 10.1104/pp.19.00405. PMID: 31253746; PMCID: PMC6752910.

37. Dasgupta, A & Klein, K. (2014). Methods for Measuring Oxidative Stress in the Laboratory. 10.1016/B978-0-12-405872-9.00002-1.

38. Shibata, T., et al. Identification of a lipid peroxidation product as a potential trigger of the p53 pathway. Journal of Biological Chemistry. 281(2), 1196-1204 (2006).

39. Skoog, D.A., West, D. M., Holler, F.J., and Crouch , S.R. Sampling, Standardization, and Calibration. In Fundamentals of Analytical Chemistry. 9th ed., Ch. 7, 153-196, Cengage-Brooks/Cole, Belmont, CA, (2014).

40. Rahman I, Li XY, Donaldson K,et. al. Glutathione homeostasis in alveolar epithelial cells in vitro and lung in vivo under oxidative stress. Am J Physiol. 1995 Sep; 269(3 Pt 1):L285-92. doi: 10.1152/ajplung.1995.269.3.L285. PMID: 7573460.

41. Cantin AM, North SL, Hubbard RC, et. al. Normal alveolar epithelial lining fluid contains high levels of glutathione. J Appl Physiol (1985). 1987 Jul; 63(1):152-7. doi: 10.1152/jappl.1987.63.1.152. PMID: 3040659.

42. D.J. Jollow, J.R. Mitchell and N. Zampaglione et al. Bromobenzene-Induced Liver Necrosis. Protective Role of Glutathione and Evidence for 3,4-Bromobenzene Oxide as the Hepatotoxic Metabolite. *Pharmacology.* Vol. 11(3):151-169. DOI: 10.1159/000136485.

43. McCord JM, Fridovich I. The reduction of cytochrome c by milk xanthine oxidase. J Biol Chem. 1968 Nov 10; 243(21):5753-60. PMID: 4972775.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

44. Marklund S, Marklund G. Involvement of the superoxide anion radical in the autoxidation of pyrogallol and a convenient assay for superoxide dismutase. Eur J Biochem. 1974 Sep 16; 47(3):469-74. doi: 10.1111/j.1432-1033.1974.tb03714.x. PMID: 4215654.

45. Misra HP, Fridovich I. The role of superoxide anion in the autoxidation of epinephrine and a simple assay for superoxide dismutase. J Biol Chem. 1972 May 25; 247(10):3170-5. PMID: 4623845.

46. Moncada S, Palmer RM, Higgs EA. Nitric oxide: physiology, pathophysiology, and pharmacology. Pharmacol Rev. 1991 Jun; 43(2):109-42. PMID: 1852778.

47. Petruzzelli S, Puntoni R, Mimotti P, Pulerà N, Baliva F, Fornai E, Giuntini C. Plasma 3-nitrotyrosine in cigarette smokers. Am J Respir Crit Care Med. 1997 Dec; 156(6):1902-7. doi: 10.1164/ajrccm.156.6.9702075. PMID: 9412573.

48. Corradi M, Majori M, Cacciani GC, Consigli GF, de'Munari E, Pesci A. Increased exhaled nitric oxide in patients with stable chronic obstructive pulmonary disease. Thorax. 1999 Jul; 54(7):572-5. doi: 10.1136/thx.54.7.572. PMID: 10377199; PMCID: PMC1745512.

49. Rutgers SR, van der Mark TW, Coers W, Moshage H, Timens W, Kauffman HF, Koëter GH, Postma DS. Markers of nitric oxide metabolism in sputum and exhaled air are not increased in chronic obstructive pulmonary disease. Thorax. 1999 Jul; 54(7):576-80. doi: 10.1136/thx.54.7.576. PMID: 10377200; PMCID: PMC1745531.

50. Maziak W, Loukides S, Culpitt S, Sullivan P, Kharitonov SA, Barnes PJ. Exhaled nitric oxide in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 1998 Mar; 157(3 Pt 1):998-1002. doi: 10.1164/ajrccm.157.3.97-05009. PMID: 9517624.

51. Arif E, Ahsan A, Vibhuti A, Rajput C, Deepak D, Athar M, Singh B, Pasha MA. Endothelial nitric oxide synthase gene variants contribute to oxidative stress in COPD.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Biochem Biophys Res Commun. 2007 Sep 14; 361(1):182-8. doi: 10.1016/j.bbrc.2007.07.008. Epub 2007 Jul 16. PMID: 17658478.

52. MacNee W. Oxidants and COPD. Curr Drug Targets Inflamm Allergy. 2005 Dec; 4(6):627-41. doi: 10.2174/156801005774912815. PMID: 17305519.

53. Hansel TT, Barnes PJ. New drugs for exacerbations of chronic obstructive pulmonary disease. Lancet. 2009 Aug 29; 374(9691):744-55. doi: 10.1016/S0140-6736(09)61342-8. PMID: 19716967.

54. Loukides S, Bakakos P, Kostikas K. Oxidative stress in patients with COPD. Curr Drug Targets. 2011 Apr; 12(4):469-77. doi: 10.2174/138945011794751573. PMID: 21194408.

55. Rahman I, MacNee W. Role of oxidants/antioxidants in smoking-induced lung diseases. Free Radic Biol Med. 1996; 21(5):669-81. doi: 10.1016/0891-5849(96)00155-4. PMID: 8891669.

56. Rahman I, MacNee W. Lung glutathione and oxidative stress: implications in cigarette smoke-induced airway disease. Am J Physiol. 1999 Dec; 277(6):L1067-88. doi: 10.1152/ajplung.1999.277.6.L1067. PMID: 10600876.

57. Rahman I, MacNee W. Role of transcription factors in inflammatory lung diseases *Thorax* 1998; **53:**601-612.

58. Church DF, Pryor WA. Free-radical chemistry of cigarette smoke and its toxicological implications. Environ Health Perspect. 1985 Dec; 64:111-26. doi: 10.1289/ehp.8564111. PMID: 3007083; PMCID: PMC1568603.

59. Nakayama T, Church DF, Pryor WA. Quantitative analysis of the hydrogen peroxide formed in aqueous cigarette tar extracts. Free Radic Biol Med. 1989; 7(1):9-15. doi: 10.1016/0891-5849(89)90094-4. PMID: 2753397.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

60. Gompertz S, Bayley DL, Hill SL, Stockley RA. Relationship between airway inflammation and the frequency of exacerbations in patients with smoking related COPD. Thorax. 2001 Jan; 56(1):36-41. doi: 10.1136/thorax.56.1.36. PMID: 11120902; PMCID: PMC1745913.

61. Uchida K, Shiraishi M, Naito Y, Torii Y, Nakamura Y, Osawa T. Activation of stress signaling pathways by the end product of lipid peroxidation. 4-hydroxy-2-nonenal is a potential inducer of intracellular peroxide production. J Biol Chem. 1999 Jan 22; 274(4):2234-42. doi: 10.1074/jbc.274.4.2234. PMID: 9890986.

62. Petruzzelli S, Puntoni R, Mimotti P, Pulerà N, Baliva F, Fornai E, Giuntini C. Plasma 3-nitrotyrosine in cigarette smokers. Am J Respir Crit Care Med. 1997 Dec; 156(6):1902-7. doi: 10.1164/ajrccm.156.6.9702075. PMID: 9412573.

63. Morrison D, Rahman I, Lannan S, MacNee W. Epithelial permeability, inflammation, and oxidant stress in the air spaces of smokers. Am J Respir Crit Care Med. 1999 Feb; 159(2):473-9. doi: 10.1164/ajrccm.159.2.9804080. PMID: 9927360.

64. Rahman I, Li XY, Donaldson K, Harrison DJ, MacNee W. Glutathione homeostasis in alveolar epithelial cells in vitro and lung in vivo under oxidative stress. Am J Physiol. 1995 Sep; 269(3 Pt 1):L285-92. doi: 10.1152/ajplung.1995.269.3.L285. PMID: 7573460.

65. Carp H, Miller F, Hoidal JR, Janoff A. Potential mechanism of emphysema: alpha 1-proteinase inhibitor recovered from lungs of cigarette smokers contains oxidized methionine and has decreased elastase inhibitory capacity. Proc Natl Acad Sci U S A. 1982 Mar; 79(6):2041-5. doi: 10.1073/pnas.79.6.2041. PMID: 6979049; PMCID: PMC346118.

66. Baraldo S, Bazzan E, Zanin ME, Turato G, Garbisa S, Maestrelli P, Papi A, Miniati M, Fabbri LM, Zuin R, Saetta M. Matrix metalloproteinase-2 protein in lung periphery is related to COPD progression. Chest. 2007 Dec 1;132(6):1733-40.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

67. Devarajan P, Johnston JJ, Ginsberg SS, Van Wart HE, Berliner N. Structure and expression of neutrophil gelatinase cDNA. Identity with type IV collagenase from HT1080 cells. J Biol Chem 1992;267: 25228-32.

68. Mott JD, Werb Z. Regulation of matrix biology by matrix metalloproteinases. Curr Opin Cell Biol 2004;16:558-64.

69. Amalinei C, Caruntu ID, Giusca SE, Balan RA. Matrix metalloproteinases involvement in pathologic conditions. Rom J Morphol Embryol. 2010;51:215-228.

70. Tacheva T, Dimov D, Vlaykova T. Gelatinases in COPD and bronchial asthma. Sci Technol Med. 2012; 2(1):255-260.

71. Fanjul-Fernandez M, Folgueras AR, Cabrera S, LopezOtin C. Matrix metalloproteinases: evolution, gene regulation and functional analysis in mouse models. Biochim Biophys Acta. 2010; 1803:3-19.

72. Visse R, Nagase H. Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. Circ Res. 2003; 92:827-839.

73. Qin H, Sun Y, Benveniste EN. The transcription factors Sp1, Sp3, and AP-2 are required for constitutive matrix metalloproteinase-2 gene expression in astroglioma cells. J Biol Chem. 1999; 274:29130-29137.

74. Tacheva, T., Chelenkova, P., Dimov, D., et. al. (2015). Frequency of the common promoter polymorphism MMP2− 1306 C>T in a population from central Bulgaria. *Biotechnology & Biotechnological Equipment*, *29* (2), 351-356.

75. Greene J, Wang M, Liu YE, et. al. Molecular cloning and characterization of human tissue inhibitor of metalloproteinase 4. J Biol Chem 1996; 271: 30375–30380.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

76. Ohnishi K, Takagi M, Kurokawa Y, Satomi S, Konttinen YJ. Matrix metalloproteinase-mediated extracellular matrix protein degradation in human pulmonary emphysema. Lab Invest 1998; 78: 1077– 1087.

77. Fischer B.M., Pavlisko E., Voynow J.A. (2011). Pathogenic triad in COPD: oxidative stress, protease-antiprotease imbalance, and inflammation. Int J Chron Obstruct Pulmon Dis, 6, 413–421.

78. Ghanei M., Ghalejooghi N.A., Nourani M.R., Harandi A.A., Fooladi A.A. (2010). Effect of TGFß1 and TIMP2 on disease activity in asthma and COPD. Iran J Allergy Asthma Immunol, 9, 2, 79–86.

79. Ramos C, Cisneros J, Gonzalez-Avila G, Becerril C, Ruiz Vc, Montaño M. Increase of Matrix Metalloproteinases in Woodsmoke-Induced Lung Emphysema in Guinea Pigs. Inhalation Toxicology 2009; 21:119-32.

80. Bakowska J, Adamson IY. Collagenase and gelatinase activities in bronchoalveolar lavage fluids during bleomycin-induced lung injury. J Pathol. 1998 Jul; 185(3):319-23. doi: 10.1002/(SICI)1096-9896(199807)185:3<319::AID-PATH114>3.0.CO;2-L. PMID: 9771487.

81. Ohnishi K, Takagi M, Kurokawa Y, Satomi S, Konttinen YT. Matrix metalloproteinase-mediated extracellular matrix protein degradation in human pulmonary emphysema. Lab Invest. 1998 Sep; 78(9):1077-87. PMID: 9759652.

82. Zhang Y, Li Y, Ye Z, Ma H. Expression of Matrix Metalloproteinase-2, Matrix Metalloproteinase-9, Tissue Inhibitor of Metalloproteinase-1, and Changes in Alveolar Septa in Patients with Chronic Obstructive Pulmonary Disease. Med Sci Monit. 2020 Oct 18; 26: e925278. Doi : 10.12659/MSM.925278. PMID: 33070147; PMCID: PMC7580176.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

83. Deb, Nilanjana & Mallick, Sudipta & Jaiswal, Ashish & Manna, Anirban & Mabalirajan, Ulaganathan & Swarnakar, Snehasikta. (2019). Role of MMPs and Oxidants in Lung Diseases. 10.1007/978-981-13-8413-4_8.

84. Ratajczak-Wielgomas K, Gosk J, Rabczyński J, Augoff K, Podhorska-Okołów M, Gamian A, Rutowski R. Expression of MMP-2, TIMP-2, TGF-β1, and decorin in Dupuytren's contracture. Connect Tissue Res. 2012; 53(6):469-77. doi: 10.3109/03008207.2012.686542. Epub 2012 Jul 24. PMID: 22512703.

85. Vitenberga-Verza, Zane & Pilmane, Māra & Babjoniseva, Aurika. (2017). COPD affected lung tissue remodelling due to the local distribution of MMP-2, TIMP-2, TGF-β1 and HSP-70. Papers on Anthropology. 26. 157. 10.12697/poa.2017.26.2.16.

86. Strongin AY, Collier I, Bannikov G, Marmer BL, Grant GA, Goldberg GI. Mechanism of cell surface activation of 72-kDa type IV collagenase. Isolation of the activated form of the membrane metalloprotease. J Biol Chem. 1995 Mar 10; 270(10):5331-8. doi: 10.1074/jbc.270.10.5331. PMID: 7890645.

87. Butler GS, Butler MJ, Atkinson SJ, Will H, Tamura T, Schade van Westrum S, Crabbe T, Clements J, d'Ortho MP, Murphy G. The TIMP2 membrane type 1 metalloproteinase "receptor" regulates the concentration and efficient activation of progelatinase A. A kinetic study. J Biol Chem. 1998 Jan 9; 273(2):871-80. doi: 10.1074/jbc.273.2.871. PMID: 9422744.

88. Peeney D, Liu Y, Lazaroff C, Gurung S, Stetler-Stevenson WG. Unravelling the distinct biological functions and potential therapeutic applications of TIMP2 in cancer. Carcinogenesis. 2022 Jun 4; 43(5):405-418. doi: 10.1093/carcin/bgac037. PMID: 35436325; PMCID: PMC9167030.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

89. Wang Z, Juttermann R, Soloway PD. TIMP-2 is required for efficient activation of proMMP-2 in vivo. J Biol Chem. 2000 Aug 25; 275(34):26411-5. doi: 10.1074/jbc.M001270200. PMID: 10827175; PMCID: PMC2683068.

90. Strongin AY, Collier I, Bannikov G, Marmer BL, Grant GA, Goldberg GI. Mechanism of cell surface activation of 72-kDa type IV collagenase. Isolation of the activated form of the membrane metalloprotease. J Biol Chem. 1995 Mar 10; 270(10):5331-8. doi: 10.1074/jbc.270.10.5331. PMID: 7890645.

91. Lähteenvuo J, Rosenzweig A. Effects of aging on angiogenesis. Circ Res. 2012 Apr 27; 110(9):1252-64. doi: 10.1161/CIRCRESAHA.111.246116. PMID: 22539758; PMCID: PMC4101916.

92. Wang, Z., Juttermann, R., and Soloway, P.D. TIMP2 is required for efficient activation of proMMP2 in vivo. J. Biol. Chem. 2000; 275, 26411-26415.

93. Tacheva T, Chelenkova P, Dimov D, Petkova R, Chakarov S, Vlaykova T. Frequency of the common promoter polymorphism MMP2− 1306 C> T in a population from central Bulgaria. Biotechnology & Biotechnological Equipment. 2015 Mar 4; 29(2):351-6.

94. Montoya J, Ramos J, Montaño M, Ramírez-Venegas A, Sansores RH, Pérez-Rubio G, Velázquez-Uncal M, Camarena A, Ramos C, Falfán-Valencia R. Genetic polymorphisms of matrix metalloproteinases and protein levels in chronic obstructive pulmonary disease in a Mexican population. Biomark Med. 2015; 9(10):979-88. doi: 10.2217/bmm.15.75. Epub 2015 Oct 6. PMID: 26439471.

95. Guo XT, Wang JF, Zhang LY, Xu GQ. Quantitative assessment of the effects of MMP-2 polymorphisms on lung carcinoma risk. Asian Pac J Cancer Prev. 2012; 13(6):2853-6. doi: 10.7314/apjcp.2012.13.6.2853. PMID: 22938472.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

96. Hirano K, Sakamoto T, Uchida Y, Morishima Y, Masuyama K, Ishii Y, Nomura A, Ohtsuka M, Sekizawa K. Tissue inhibitor of metalloproteinases-; 2 gene polymorphisms in chronic obstructive pulmonary disease. European Respiratory Journal. 2001 Nov 1; 18(5):748-52.

97. Hegab AE, Sakamoto T, Uchida Y, Nomura A, Ishii Y, Morishima Y, Mochizuki M, Kimura T, Saitoh W, Kiwamoto T, Iizuka T. Association analysis of tissue inhibitor of metalloproteinase2 gene polymorphisms with COPD in Egyptians. Respiratory medicine. 2005 Jan 1; 99(1):107-10.

98. Korytina GF, Akhmadishina LZ, Viktorova EV, Tselousova OS, Danilko KV, Kochetova OV, Viktorova TV. Extracellular matrix remodeling genes polymorphisms and risk of chronic bronchitis and recurrent pneumonia in children. J Hum Genet. 2013 Jul; 58(7):467-74. doi: 10.1038/jhg.2013.24. Epub 2013 Apr 11. PMID: 23575435.

99. Global Initiative for Chronic Obstructive Lung Disease (GOLD): GOLD 2017 global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease, 2017 report. [https://goldcopd.org/gold-2017-global-strategy-diagnosis-management-%20prevention-copd/]

100. Koul P A. Chronic obstructive pulmonary disease: Indian guidelines and the road ahead. Lung India 2013; 30:175-7.

101. Gueders M M, Foidart J M, Noel A, Cataldo D D. Matrix metalloproteinases (MMPs) and tissue inhibitors of MMPs in the respiratory tract: potential implications in asthma and other lung diseases. Eur J Pharmacol 2006; 533:133-44.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

102. Ramos C, Cisneros J, Gonzalez-Avila G, Becerril C, Ruiz Vc, Montaño M. Increase of Matrix Metalloproteinases in Woodsmoke-Induced Lung Emphysema in Guinea Pigs. Inhalation Toxicology 2009; 21:119-32.

103. Wills-Karp M. Interleukin-13 in asthma pathogenesis. Immunol Rev 2004; 202:175-90.

104. Owen C A, Hu Z, Barrick B, Shapiro S D. Inducible expression of tissue inhibitor of metalloproteinases-resistant matrix metalloproteinase-9 on the cell surface of neutrophils. Am J Respir Cell Mol Biol 2003; 29:283-94.

105. Culpitt S V, Rogers D F, Traves S L, Barnes P J, Donnelly L E. Sputum matrix metalloproteases: comparison between chronic obstructive pulmonary disease and asthma. Respiratory Medicine 2005; 99:703-10.

106. Docherty A J, Lyons A, Smith B J, Wright E M, Stephens P E, Harris T J et al. Sequence of human tissue inhibitor of metalloproteinases and its identity to erythroid-potentiating activity. Nature 1985; 318:66-9.

107. Hirano K, Sakamoto T, Uchida Y, Morishima Y, Masuyama K, Ishii Y, Nomura A, Ohtsuka M, Sekizawa K. Tissue inhibitor of metalloproteinases-2 gene polymorphisms in chronic obstructive pulmonary disease. Eur Respir J. 2001 Nov; 18(5):748-52. doi: 10.1183/09031936.01.00102101. PMID: 11757622.

108. Barnes PJ. Chronic Obstructive Pulmonary Disease. N Engl J Med 2000; 343:269-280. DOI: 10.1056/NEJM20000727343040.

109. Hegab AE, Sakamoto T, Uchida Y, Nomura A, Ishii Y, Morishima Y, Mochizuki M, Kimura T, Saitoh W, Kiwamoto T, Iizuka T, Massoud HH, Massoud HM, Hassanein KM, Sekizawa K. Association analysis of tissue inhibitor of metalloproteinase2 gene

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

polymorphisms with COPD in Egyptians. Respir Med. 2005 Jan; 99(1):107-10. doi: 10.1016/j.rmed.2004.05.017. PMID: 15672858.

110. Tanya Tacheva, Dimo Dimov, Pavlina Chelenkova, Rumena Petkova, Krasimir D imov, Stoyan Chakarov, Tatyana Vlaykova. Eur Respir J. 2015 46: PA861; **DOI:** 10.1183/13993003.congress-2015.PA861.

111. Ishida M, Mikami S, Kikuchi E, Kosaka T, Miyajima A, Nakagawa K, Mukai M, Okada Y, Oya M. Activation of the aryl hydrocarbon receptor pathway enhances cancer cell invasion by upregulating the MMP expression and is associated with poor prognosis in upper urinary tract urothelial cancer. Carcinogenesis. 2010 Feb 1; 31(2):287-95.

112. John A, Tuszynski G. The role of matrix metalloproteinases in tumor angiogenesis and tumor metastasis. Pathol Oncol Res. 2001; 7(1):14-23. doi: 10.1007/BF03032599. PMID: 11349215.

113. Ala-aho R, Kähäri VM. Collagenases in cancer. Biochimie. 2005 Mar-Apr; 87(3-4):273-86. doi: 10.1016/j.biochi.2004.12.009. PMID: 15781314.

114. Ye S. Polymorphism in matrix metalloproteinase gene promoters: implication in regulation of gene expression and susceptibility of various diseases. Matrix Biol. 2000 Dec; 19(7):623-9. doi: 10.1016/s0945-053x(00)00102-5. PMID: 11102751.

115. H.E. Van Wart, H. Birkedal-Hansen. The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. Proc. Natl. Acad. Sci. USA, 87 (1990), pp. 5578-5582.

116. S. Rajagopalan, X.P. Meng, S. Ramasamy, D.G. Harrison, Z.S. Galis. Reactive oxygen species produced by macrophage-derived foam cells regulate the activity of vascular matrix metalloproteinases in vitro: implications for atherosclerotic plaque stability. J. Clin. Invest., 98 (1996), pp. 2572-2579.

117. F. Valentin, J.L. Bueb, P. Kieffer, E. Tschirhart, J. Atkinson. Oxidative stress activates MMP-2 in cultured human coronary smooth muscle cells. Fundam. Clin. Pharmacol., 19 (2005), pp. 661-667.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

118. K. Grote, I. Flach, M. Luchtefeld, E. Akin, S.M. Holland, H. Drexler, B. Schieffer . Mechanical stretch enhances mRNA expression and proenzyme release of matrix metalloproteinase-2 (MMP-2) via NAD(P)H oxidase-derived reactive oxygen species. Circ. Res., 92 (2003), pp. e80-e86.

119. R. Schulz. Intracellular targets of matrix metalloproteinase-2 in cardiac disease: rationale and therapeutic approaches. Annu. Rev. Pharmacol. Toxicol., 47 (2007), pp. 211-242.

120. S. Viappiani, A.C. Nicolescu, A. Holt, G. Sawicki, B.D. Crawford, H. Leon, T. va n Mulligen, R. Schulz. Activation and modulation of 72 kDa matrix metalloproteinase-2 by peroxynitrite and glutathione. Biochem. Pharmacol., 77 (2008), pp. 826-834.

121. Zhang HJ, Zhao W, Venkataraman S, Robbins ME, Buettner GR, Kregel KC, Oberley LW: Activation of matrix metalloproteinase-2 by overexpression of manganese superoxide dismutase in human breast cancer MCF-7 cells involves reactive oxygen species. *J Biol Chem* 2002, 277(23):20919-20926. 10.1074/jbc.M109801200.

122. ManDal, M., Das, S., Chakraborti, T. *et al.* Role of matrix metalloprotease-2 in oxidant activation of Ca^{2+}ATPase by hydrogen peroxide in pulmonary vascular smooth muscle plasma membrane. *J Biosci* **28**, 205–213 (2003). https://doi.org/10.1007/BF02706220.

123. Kawaguchi, Y., Tanaka, H., Okada, T. *et al.* The effects of ultraviolet A and reactive oxygen species on the mRNA expression of 72-kDa type IV collagenase and its tissue inhibitor in cultured human dermal fibroblasts. *Arch Dermatol Res* **288**, 39–44 (1996). https://doi.org/10.1007/BF02505041.

124. Hashimoto T, Uchida K, Okayama N, Imate Y, Suehiro Y, Hamanaka Y, Ueyama Y, Shinozaki F, Yamashita H, Hinoda Y. Association of matrix metalloproteinase (MMP)-1 promoter polymorphism with head and neck squamous cell carcinoma, Cancer Letters, Vol 211(1), 2004,19-24.

125. Van Diemen, C., Postma, D., Siedlinski, M. *et al.* Genetic variation in *TIMP1* but not *MMPs* predict excess FEV_1 decline in two general population-based cohorts. *Respir Res* **12**, 57 (2011). https://doi.org/10.1186/1465-9921-12-57.

126. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, van Weel C, Zielinski J; Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

2007 Sep 15; 176(6):532-55. doi: 10.1164/rccm.200703-456SO. Epub 2007 May 16. PMID: 17507545.

127. Kim CH, Oh Y, Lee TH. Codon optimization for high-level expression of human erythropoietin (EPO) in mammalian cells. Gene. 1997 Oct 15; 199(1-2):293-301. doi: 10.1016/s0378-1119(97)00384-3. PMID: 9358069.

128. Ross J. mRNA stability in mammalian cells. Microbiol Rev. 1995 Sep; 59(3):423-50. doi: 10.1128/mr.59.3.423-450.1995. PMID: 7565413; PMCID: PMC239368.

129. Ohnishi K, Takagi M, Kurokawa Y, Satomi S, Konttinen YT. Matrix metalloproteinase-mediated extracellular matrix protein degradation in human pulmonary emphysema. Lab Invest. 1998 Sep; 78(9):1077-87. PMID: 9759652.

130. De Clerck YA, Darville MI, Eeckhout Y, Rousseau GG. Characterization of the promoter of the gene encoding human tissue inhibitor of metalloproteinases-2 (TIMP-2). Gene. 1994 Feb 25; 139(2):185-91. doi: 10.1016/0378-1119(94)90753-6. PMID: 8112602.

131. Willemse BWM, ten Hacken NHT, Rutgers B, Lesman-Leegte I, Postma DS, Timens W. Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers. European Respiratory Journal. 2005 Nov; 26(5):835-45.

132. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med. 2013 May 15; 187(10):1067-75.

133. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. Science. 2004 Mar 5; 303(5663):1532-5.

134. Yipp BG, Kubes P. NETosis: how vital is it? Blood. 2013 Oct 17;122(16):2784-94

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

135. Papayannopoulos V, Metzler KD, Hakkim A, Zychlinsky A. Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps. J Cell Biol. 2010 Nov 1; 191(3):677-91.

136. Grabcanovic-Musija F, Obermayer A, Stoiber W, Krautgartner WD, Steinbacher P, Winterberg N, et al. Neutrophil extracellular trap (NET) formation characterises stable and exacerbated COPD and correlates with airflow limitation. Respir Res. 2015; 16:59.

137. Obermayer A, Stoiber W, Krautgartner W-D, Klappacher M, Kofler B, Steinbacher P, et al. New Aspects on the Structure of Neutrophil Extracellular Traps from Chronic Obstructive Pulmonary Disease and In Vitro Generation. Plos One. 2014 15/05/2014; 9(5).

138. Pedersen F, Marwitz S, Holz O, Kirsten A, Bahmer T, Waschki B, et al. Neutrophil extracellular trap formation and extracellular DNA in sputum of stable COPD patients. Respir Med. 2015 Oct; 109(10):1360-2.

139. Foell D, Wittkowski H, Vogl T, Roth J. S100 proteins expressed in phagocytes: a novel group of damage-associated molecular pattern molecules. J Leukoc Biol. 2007; 81:28–37.

140. Crooks SW, Bayley DL, Hill SL, et al. Bronchial inflammation in acute bacterial exacerbations of chronic bronchitis: the role of leukotriene B4. Eur Respir J 2000; 15: 274–280.

141. Haas TL, Stitelman D, Davis SJ, Apte SS, Madri JA. Egr-1 mediates extracellular matrix-driven transcription of membrane type 1 matrix metalloproteinase in endothelium. J Biol Chem 1999; 274:22679–22685.

142. Ning W, Li CJ, Kaminski N, Feghali-Bostwick CA, Alber SM, Di YP, Otterbein SL, Song R, Hayashi S, Zhou Z, Pinsky DJ, Watkins SC, Pilewski JM, Sciurba FC, Peters

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

Pre-Proof

DG, Hogg JC, Choi AM. Comprehensive gene expression profiles reveal pathways related to the pathogenesis of chronic obstructive pulmonary disease. Proc Natl Acad Sci USA 2004; 101:14895–14900.

143. Lappalainen U, Whitsett JA, Wert SE, Tichelaar JW, Bry K. Interleukin-1beta causes pulmonary inflammation, emphysema, and airway remodeling in the adult murine lung. Am J Respir Cell Mol Biol 2005; 32:311– 318.

144. Corry DB, Rishi K, Kanellis J, Kiss A, Song Lz LZ, Xu J, Feng L, Werb Z, Kheradmand F. Decreased allergic lung inflammatory cell egression and increased susceptibility to asphyxiation in MMP2- deficiency. Nat Immunol 2002; 3:347–353.

145. Fink MP. Role of reactive oxygen and nitrogen species in acute respiratory distress syndrome. Curr Opin Crit Care 2002; 8:6–11.

146. Finkel T, Holbrook NJ. Oxidants, oxidative stress and the biology of ageing. Nature 2000; 408:239– 247.

147. Lass A, Suessenbacher A, Wolkart G, Mayer B, Brunner F. Functional and analytical evidence for scavenging of oxygen radicals by L-arginine. Mol Pharmacol 2002; 61:1081–1088.

148. Nicoglou A, Merlin F. Epigenetics: A way to bridge the gap between biological fields. Stud Hist Philos Biol Biomed Sci. 2017 Dec; 66:73-82. doi: 10.1016/j.shpsc.2017.10.002. Epub 2017 Oct 13. PMID: 29033228.

149. Rahman I, Marwick J, Kirkham P. Redox modulation of chromatin remodeling: impact on histone acetylation and deacetylation, NF-kB and pro-inflammatory gene expression. Biochem Pharmacol. 2004 Sep 15; 68(6):1255-67. doi: 10.1016/j.bcp.2004.05.042. PMID: 15313424.

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

150. Marwick JA, Kirkham PA, Stevenson CS, Danahay H, Giddings J, Butler K, Donaldson K, Macnee W, Rahman I. Cigarette smoke alters chromatin remodeling and induces proinflammatory genes in rat lungs. Am J Respir Cell Mol Biol. 2004 Dec; 31(6):633-42. doi: 10.1165/rcmb.2004-0006OC. Epub 2004 Aug 27. PMID: 15333327.

151. Hoekstra R, Eskens FA, Verweij J. Matrix metalloproteinase inhibitors: current developments and future perspectives. Oncologist. 2001; 6(5):415-27. doi: 10.1634/theoncologist.6-5-415. PMID: 11675519.

152. Falcone D, Gallelli L, Di Virgilio A, Tucci L, Scaramuzzino M, Terracciano R, Pelaia G, Savino R. Effects of simvastatin and rosuvastatin on RAS protein, matrix metalloproteinases and NF-κB in lung cancer and in normal pulmonary tissues. Cell Prolif. 2013 Apr; 46(2):172-82. doi: 10.1111/cpr.12018. PMID: 23510472; PMCID: PMC6495807.

Pre-P

Copyright *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation* ©2023
Published online August 3, 2023 https://doi.org/10.15326/jcopdf.2022.0370

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Sarcasm Detection on News Headline Dataset Using Language Models

Abhilasha Sharma
Software Engineering
Delhi Technological University
New Delhi, India
abhi16.sharma@gmail.com

Abhi Uday Pandey
Software Engineering
Delhi Technological University
New Delhi, India
abhiuday102@gmail.com

Apoorv Gupta
Software Engineering
Delhi Technological University
New Delhi, India
guptaapoorv0412@gmail.com

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205839

***Abstract*—Mechanized detection of sarcasm in textual data is a challenging task in NLP, as it involves recognizing the opposite of what is actually meant. This has become important today since many artificial intelligence tasks require understanding the user and answering appropriately. In this paper, we investigate multiple deep learning models for sarcasm detection on a news headline dataset, including BiLSTM and BERT. For the first time, hybrid BERT models have been used in our study to identify sarcasm in non-social media writings. The results show that the BERT and GBDT strategy combined performs better than all the other strategies listed.**

***Index Terms*—Sarcasm Detection, Natural Language Processing, BERT, Bi-directional LSTM, Support Vector Classifier, Gradient Boosted Decision Tree, Sarcasm,**

I. Introduction

Sarcasm is referred to an expression where the person uses it to say the opposite of what they mean, either to criticize something or for humor.

Unlike tweets which have a context and hence it becomes easier to detect sarcasm in them, sarcastic news headlines which have single line are much more difficult to detect and understand. This becomes an important use case for tools and artificial intelligences which have to summarize news articles for business use-case since sarcastic headlines if understood as non-sarcastic can completely change the meaning of the text. A sarcastic headline could also be used to target a certain, race, or ethnicity. To prevent this one can prevent propagation of such content which can cause major troubles. We attempt to perform this task using a various deep learning models and a combination of such techniques to reduce computation requirements and increase accuracy at the same time. We have attempted to solve their short-comings and improve considerably on the existing work [10].

Detecting Sarcasm is a complicated task for humans let alone for machines due to its dependency on the speaker's perception and knowledge of related matters such as culture, language, surrounding news, etc. Despite its challenges, the automatic detection of sarcasm is very important for the identification of hateful texts and comments that could be inappropriate for some sections of society. Detecting this sarcasm in headlines of news and social media posts becomes more difficult due to lack of context and limited information. To improve upon the previous works of sarcasm detection this study explores various language models such as Bi-LSTM (Bidirectional Long Short Term Memory) and BERT (Bidirectional Encoder Representations from Transformers) to tackle issues faced by the previous work. This study also proposes 2 hybrid architectures to optimize the results with the limited hardware resources utilising a combination of BERT with other techniques like BERT+SVC (Support Vector Classifier) and BERT+GBDT (Gradient Boost Decision Trees) we can achieve decent results on this dataset.

The involved dataset[4], [5], [31] is the News Headlines Dataset for Sarcasm Detection which contains news headlines from two journal websites. This dataset after checking for any bias we train the models and attempt to achieve better accuracy than previous works and with limited resources.

II. Previous Works

The subjectivity in detecting sarcasm in a piece of text makes it a very difficult NLP task. It boils down to a classification task where we attempt to predict whether the input text is sarcastic or not. Here we discuss the main approaches that have addressed this task using all types of datasets.

A. Traditional Approaches

One of the early approaches Liebrecht et.al [8] presents an approach based on a balanced Winnow that performed text classification. This generates results according to weight classes, these weights are then used to discover and segregate the features in a sorted format for a specific class. They explored the experiment in both formats, when the data was balanced and when it was unbalanced. The dataset comprised various tweets where the sarcastic ones contained hashtags The classifier provided a 75% True Positive Rate and a 16% False Positive Rate when working on the balanced dataset. According to the authors, the use of markers such as hashtags in digital communication can be likened to nonverbal expressions, and their inclusion can signal a sarcastic tone.

In 2013, Riloff et al. [10] provided a different method to recognize sarcasm in the text.. The used key-point here was that sarcasm is referring to the opposite of what you mean. Their approach works by identifying the type of contrast in the text between positive and negative sentiments present in the sentence This was accomplished using a cutting-edge

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:34 UTC from IEEE Xplore. Restrictions apply.

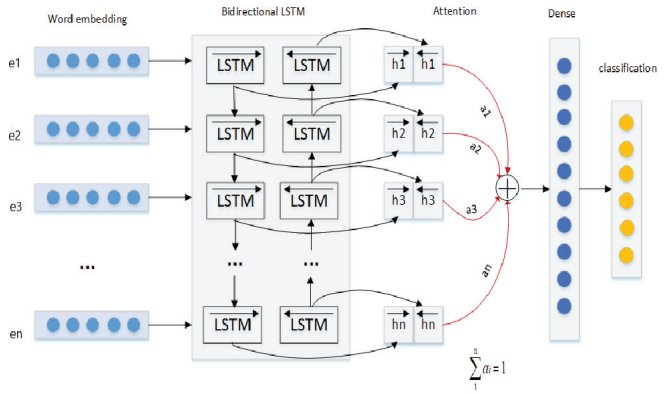


Fig. 1. General Bi-LSTM model architecture Source: retrieved from [32]

bootstrap algorithm that, when present, distinguishes tweets' positive and negative sentiments. After that, positive and negative terms were searched for in the tweets in order to identify sarcasm using bootstrapped lexicons. The F1 score for this approach was 22. The hybrid system's score was increased to 51 by combining the aforementioned with a support vector machine.

Another strategy involved making an effort to classify the comments in the context of the references. This is so that the intended irony can be understood, and to do so, additional information about the tweet was integrated with the Twitter data. The features of this additional information included things like the author's name, details about the user profiles, people who were involved, related news, etc. Bamman et al[10] made this suggestion in 2015, and it had an accuracy rate of 85%.

We attempt to combine the good parts of the above-mentioned approaches to get better accuracy such as hybrid network and using context-free data since it's a more challenging task.

B. Deep Learning Approaches

Amir et al [2] presented a technique to identify sarcasm by using a combination of embeddings with lexical features. These embeddings provides the relation between the users and the content they provide. This reduces the work since it avoids the complicated feature engineering process. Using a Convulational Neural network an accuracy of 87% was achieved on the training set, however the test set accuracy was not very encouraging.

More recent studies have experimented with BERT due to its good results in other Natural Language tasks. Xu et al.(2019)[6] conducted an investigation on different models to understand the effect of context on sarcasm. This is one of the places BERT is utilized. The BERT although a very good model is not suitable for data which has limited context like the news headline dataset. This particular study utilised a LSTM model which provided highest accuracy of 73% for a discussion forum dataset.

Khatri & P (2020)[3] also suggested utilizing machine learning classifiers in conjunction with BERT and GloVe embeddings to identify sarcasm in tweets. In this method, various classifiers like Random Forest, Logistic Regression, Support Vectot Machines, and Naive Bayes were tested. The above study proved embeddings to be a powerful tool for this task, since they convert the text into corresponding vector representations

The main learnings from the previous work are

1. Usage of embeddings increases accuracy
2. Using a Bi-directional model is beneficial since the words ahead often determine the context of the words coming first in the sentence
3. A hybrid architecture is the best solution as all models have some flaws in them.

III. Methodology

A. Baseline Model

NAÏVE BAYES CLASSIFIER: The baseline Classifier here is Naive Bayes Classifier [34] This classifier is used so as to get the probabilistic mean for sarcastic headlines in the text features dataset. It is applied after applying a count vectorizer with the analyzer = word and tf-idf transformer with the smooth_idf = True on the dataset.

This is a popular probabilistic machine-learning approach for classification problems. Based on the Bayes theorem, a mathematical formula expressing the likelihood of an event provided another event has already occurred, is the foundation of the algorithm. Given the values of the features connected to a given data point, the Naive Bayes Classifier uses Bayes' theorem to determine the likelihood that the data point belongs to a specific class. Using the prior work as a guide, we choose the model that provides the best test accuracy and enhance it using this as the baseline model.

Test accuracy 79.5%

Note a nearly 80% accuracy is still decent since detection of sarcasm is sometimes difficult for humans as well let alone of machines, hence this acts as a good starting point upon which we can improve the results.

B. Proposed Models

Here In total 4 models are proposed Namely, Bi-LSTM, BERT, BERT+SVC and BERT+GBDT.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:34 UTC from IEEE Xplore. Restrictions apply.

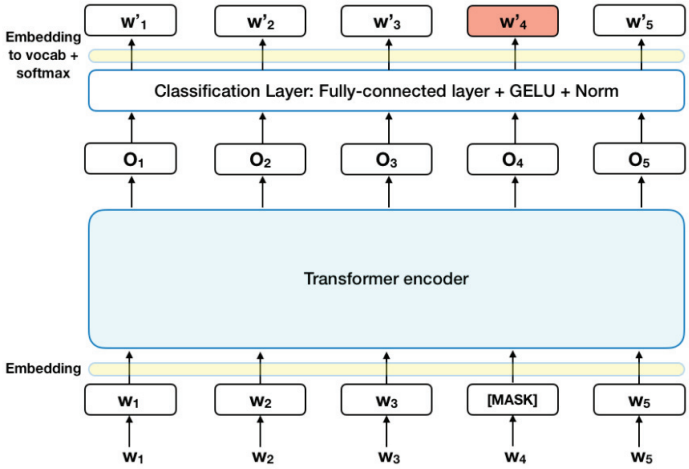


Fig. 2. General BERT model architecture Source: retrieved from [33]

C. Bi-LSTM

Bi-LSTM (Bidirectional Long Short-Term Memory) is a modification on the traditional recurrent neural network that can been used in sarcasm detection. Bi-LSTMs have the ability to depict unordered and long-term dependencies in sequential data, making them well-suited for detecting sarcasm in text.

Advantages of using Bi-LSTM for sarcasm detection include:

1. Capturing long-term dependencies: Bi-LSTMs have a memory cell that can store information for a longer time period. This proves useful in capturing long-term dependencies in sequential data, making them effective for detecting sarcasm that may depend on the previous context.
2. Less computational requirements: Bi-LSTMs require less computational resources compared to models like BERT, making them more accessible for users with limited resources. The model parameters can be seen in Figure 3.

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 400, 300)          6000000

 conv1d (Conv1D)             (None, 398, 32)           28832

 max_pooling1d (MaxPooling1D  (None, 199, 32)          0
 )

 dropout (Dropout)           (None, 199, 32)           0

 bidirectional (Bidirectiona  (None, 199, 240)         146880
 l)

 flatten (Flatten)           (None, 47760)             0

 dense (Dense)               (None, 35)                1671635

 dropout_1 (Dropout)         (None, 35)                0

 dense_1 (Dense)             (None, 2)                 72

=================================================================
Total params: 7,847,419
Trainable params: 1,847,419
Non-trainable params: 6,000,000
_________________________________________________________________
```

Fig. 3. Bi-LSTM Model Used in the Study

D. BERT

BERT is an acronym for Bidirectional Encoder Representations from Transformers, it is a pre-trained language model based on transformer architecture. It has been trained on a massive amount of text data using a masked language modeling approach, where it learns to predict masked words in a sentence based on its context.

One of the significant advantages of BERT is its ability to capture context and meaning in a sentence, making it particularly useful for natural language processing tasks such as text classification, question answering, and extracting the sentiment or emotion expressed in a piece of text BERT's architecture includes multiple layers of transformers, which enable it to understand the context of a word within a sentence and even across multiple sentences.

The model contains 12 layers of transformer block with a hidden size of 768 the L = 12 and H = 768, The parameters for the same can be found in Figure 4.

```
bert_arch(
  (bert): BertModel(
    (embeddings): BertEmbeddings(
      (word_embeddings): Embedding(30522, 768, padding_idx=0)
      (position_embeddings): Embedding(512, 768)
      (token_type_embeddings): Embedding(2, 768)
      (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)
      (dropout): Dropout(p=0.1, inplace=False)
    )
```

Fig. 4. BERT Model Used in the study

E. BERT+SVC

Bert+SVC is a machine learning technique that combines the power of two different algorithms: BERT and Support Vector Machines (SVC). BERT has achieved great results on a wide range of natural language processing (NLP) tasks. Support Vector Machines, on the other hand, are a powerful class of algorithms for binary and multiclass classification. The BERT model is trained on a substantial volume of textual data and is able to generate high-quality word embeddings that

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:34 UTC from IEEE Xplore. Restrictions apply.

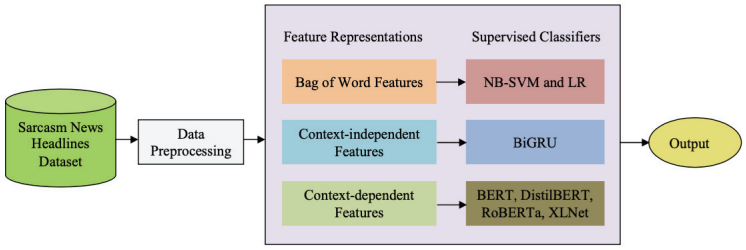


Fig. 5. A general framework for supervised learning models to detect sarcasm. Source: retrieved from [31]

capture the semantic and contextual meaning of words. These embeddings can be used to train a classifier for a specific NLP task, such as sentiment analysis or named entity recognition. However, the BERT model alone may not always be able to capture all of the nuances of a specific task, especially if the training data is limited. This is where SVC comes in. SVC is a type of binary classifier that is trained to maximize the margin between two classes of data points. This means that it is able to identify complex decision boundaries that can accurately classify data points in high-dimensional spaces. For creating text embedding, we have used the Bert model, and for processing images, we have used the SVC model. When combined with the BERT model, the SVC classifier can learn to recognize subtle patterns in the data that may be missed by BERT alone.

F. BERT+GBDT

BERT+GBDT is a machine learning technique that combines the power of BERT and Gradient Boosting Decision Trees (GBDT) for sarcasm detection. The BERT model alone sometimes is unable to capture all the features required to determine whether the sentence is sarcastic or not This is where GBDT comes in. GBDT is a machine learning algorithm that builds an ensemble of decision trees to classify data points. It works by sequentially adding decision trees that correct the errors of the previous trees, resulting in a highly accurate classifier. When combined with the BERT model, the GBDT classifier can learn to recognize subtle patterns in the data that may be missed by BERT alone. Here the sentences are first preprocessed and tokenized using BERT. The GBDT classifier is then inputted or fed with the resulting word embeddings, which learns to distinguish between sarcastic and non-sarcastic sentences. Therefore BERT acts as an interpreter for the Gradient boost method.

IV. Dataset

The dataset of news headlines, benchmarked by the authors, is utilized for sarcasm detection, with two versions available: version 1 and version 2. Version 1 includes a total of 26,709 news headlines, comprising 11,724 sarcastic and 14,985 non-sarcastic headlines. On the other hand, version 2 consists of 28,619 records, with 13,634 sarcastic and 14,985 non-sarcastic headlines, including an additional 1910 headlines from version 1. Due to the larger number of headlines, version 2 is preferred for use and used in this study

A. Properties

This dataset contains 3 columns
1. Headline
2. The link of the news piece
3. 0 or 1 value where 1 means the headline is sarcastic and 0 implies it's not sarcastic
The 3rd column acts as its label for supervised learning The dataset contains 28619 records where 14985 are non-sarcastic while others (i.e. 13634) are sarcastic.

B. Advantages over twitter dataset

1. Since the headlines are composed by professionals, the sentences lack gramatical errors and spelling errors. This increases the chances of better fit with existing embeddings without loss of any important words.
2. The target website involved are specifically made saracastic so there is little chance of mislabeling.
Headlines alone are often context-free and hence easier for the model to understand since some context cannot be understood or known to the machine.
3.. For instance, a sarcastic tweet about a recent fire at new york museum will be very hard to understand for people who are unaware about such an incident.

V. Result and Conclusion

A. Evaluation Metric

As used in most supervised learning algorithms and in the base model the accuracy of the model is determined by the ratio of headlines that are classified correctly to that of total headlines in the data set.

$$Accuracy = \frac{Number\ of\ headlines\ classified\ correctly}{Total\ numbers\ of\ headlines\ in\ the\ set} * 100$$

This accuracy is considered for the test data formed from the 80-20 split pre-processing libraries. This split is the same for all the models. We have used only the test accuracy since some models tend to over-fit on the data producing very high accuracy on the training set which is often a misleading metric to judge the model

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:34 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARATIVE ANALYSIS

Model	Test Accuracy
XGBOOST CLASSIFIER	**67.5**
LOGISTIC REGRESSION CLASSIFIER	**77.5**
NAIVE BAYES (Baseline)	**79.5**
Bi-LSTM	**87.5**
BERT	**64.74**
BERT + SVC	**85.17**
BERT + GBDT	**91.48**

B. Results achieved

The above-mentioned models were executed in a Google Colab Pro environment using standard libraries such as nltk for pre-processing and embeddings. To feed the textual data was first converted to machine-readable format using word embeddings. After performing standard pre-processing the data was split in an 80-20 form

A comparative analysis among the 4 supervised learning algorithms namely Bi-LSTM, BERT, BERT+SVC, and BERT+GBDT. shows BERT+GBDT achieved the highest test accuracy. The Bi-LSTM Model aids in taking into account the context of previous words from both sides which becomes crucial for sentences whose meaning is determined by the ending words. Here we have used 300d glove embeddings as opposed to the predecessors which utilised 50d or 200d embeddings. Also, we were able to achieve a better results by tweeking few model parameters such as the dropout layer and neurons in the network. The optimum number was achieved through rigourous hit and trial.

The BERT has high GPU requirements when trained by unlocking the layers, this is a major drawback of using BERT for this task due to the high number of entries (nearly 29,000). Based on previous work with BERT and our own results we concluded that BERT is very good in gathering the details and pattern in the data however, it's training for the same was not very practical with the hardware resources at our disposal, we, therefore, utilised the best feature of BERT of pattern recognition and formed a completely new architecture which uses an SVC classifier. We were further able to increase accuracy by combining Gradient boost decision trees with BERT This hybrid architecture gives the best accuracy for the dataset and it also requires less computation time.

The usage of 300d glove embeddings and a Bi-directional model increases the existing accuracy of Bi-LSTM Based on the context-less nature of headlines the proposed models perform considerably well and better than the baseline model The results of each model are summarised in Table 1

REFERENCES

[1] S. S. Hossain, Y. Arafat, and M. E. Hossain, "Context-based news headlines analysis: A comparative study of machine learning and deep learning algorithms," Vietnam Journal of Computer Science, pp. 1–15, 2021.

[2] Amir, Silvio & Wallace, Byron & Lyu, Hao & Carvalho, Paula & Silva, Mário. (2016). Modelling Context with User Embeddings for Sarcasm Detection in Social Media. 167-177. 10.18653/v1/K16-1017.

[3] [Sarcasm Detection in Tweets with BERT and GloVe Embeddings](https://aclanthology.org/2020.figlang-1.7) (Khatri & P, Fig-Lang 2020)

[4] Misra, Rishabh and Prahal Arora. "Sarcasm Detection using News Headlines Dataset." AI Open (2023).

[5] Misra, Rishabh and Jigyasa Grover. "Sculpting Data for ML: The first act of Machine Learning." ISBN 978-0-578-83125-1 (2021).

[6] [BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis](https://aclanthology.org/N19-1242) (Xu et al., NAACL 2019)

[7] Wang, Zhe, Rundong Shi and Shijie Li. "GBDT and BERT: a Hybrid Solution for Recognizing Citation Intent." (2020).

[8] Liebrecht, Christine & Kunneman, Florian & Van den Bosch, Antal. (2013). The perfect solution for detecting sarcasm in tweets #not. 29-37.

[9] Saha, Shubhodip & Yadav, Jainath & Ranjan, Prabhat. (2017). Proposed Approach for Sarcasm Detection in Twitter. Indian Journal of Science and Technology. 10. 10.17485/ijst/2017/v10i25/114443.

[10] Scola, E., & Segura-Bedmar, I. (2021). Sarcasm Detection with BERT. Procesamiento Del Lenguaje Natural, 67, 13-25. Retrieved from http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6373

[11] S. Poria E. Cambria D. Hazarika and P. Vij "A deeper look into sarcastic tweets using deep convolutional neural networks " 2016.

[12] I. Chaturvedi, Y.-S. Ong, I. Tsang, R. Welsch, and E. Cambria, "Learning word dependencies in text by means of a deep recurrent belief network," Knowledge-Based Systems, vol. 108, pp. 144–154, 2016.

[13] S. Kumar, M. Gahalawat, P. P. Roy, D. P. Dogra, and B.-G. Kim, "Exploring impact of age and gender on sentiment analysis using machine learning," Electronics, vol. 9, no. 2, p. 374, 2020.

[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018.

[15] M. Zhang, Y. Zhang, and G. Fu, "Tweet sarcasm detection using deep neural network," in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: technical papers, 2016, pp. 2449–2460.

[16] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," Advances in neural information processing systems, vol. 32, 2019.

[17] A. Joshi, V. Sharma, and P. Bhattacharyya, "Harnessing context incongruity for sarcasm detection," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015, pp. 757–762.

[18] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[20] https://towardsdatascience.com/sarcasm-detection-with-nlp-cbff1723f69a

[21] T. Jain, N. Agrawal, G. Goyal and N. Aggrawal, "Sarcasm detection of tweets: A comparative study," 2017 Tenth International Conference on Contemporary Computing (IC3), Noida, India, 2017, pp. 1-6, doi: 10.1109/IC3.2017.8284317.

[22] Ashwitha A, Shruthi G, Shruthi H R, Makarand Upadhyaya, Abhra Pratip Ray, Manjunath T C, Sarcasm detection in natural language processing, Materials Today: Proceedings,Volume 37, Part 2, 2021,

[23] S.Amir,B.C.Wallace,H.Lyu,P.Carvalho,andM.J.Silva,"Modelling context with user embeddings for sarcasm detection in social media," in Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning, 2016, pp. 167–177.

[24] [Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm](https://aclanthology.org/D17-1169) (Felbo et al., EMNLP 2017)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:34 UTC from IEEE Xplore. Restrictions apply.

[25] Abigail See, Peter J. Liu, Christopher D. Manning. (2017). Get To The Point: Summarization with Pointer-Generator Networks.
[26] Jingqiang Chen, Hai Zhuge,Extractive summarization of documents with images based on multi-modal RNN, Future Generation Computer Systems, Volume 99, 2019
[27] A. Joshi, P. Bhattacharyya, and M. J. Carman, "Automatic sarcasm detection: A survey," ACM Computing Surveys (CSUR), vol. 50, no. 5, pp. 1–22, 2017.
[28] S. Mukherjee and P. K. Bala, "Sarcasm detection in microblogs using na ıve bayes and fuzzy clustering," Technology in Society, vol. 48, pp. 19–27, 2017.
[29] H. Nayel, E. Amer, A. Allam, and H. Abdallah, "Machine learning-based model for sentiment and sarcasm detection," in Proceedings of the Sixth Arabic Natural Language Processing Workshop, 2021, pp. 386–389.
[30] A. K. Jena, A. Sinha, and R. Agarwal, "C-net: Contextual network for sarcasm detection," in Proceedings of the Second Workshop on Figurative Language Processing, 2020, pp. 61–66.
[31] https://www.researchgate.net/figure/Sarcasm-detection-LSTM-architecture
[32] https://www.researchgate.net/figure/The-architecture-of-BiLSTM-Attention
[33] https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270
[34] https://github.com/bhomikpathak02/Sarcasm-detection-in-news-headlines
[35] https://www.kaggle.com/rmisra/news-headlines-dataset-for-sarcasm-detection
[36] https://towardsdatascience.com/how-to-turn-text-into-features-478b57632e99

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:19:34 UTC from IEEE Xplore. Restrictions apply.

SDL-Net: A combined CNN & RNN Human Activity Recognition model

Dhruv Gupta*
Biometric Research Laboratory
Delhi Technological University,
New Delhi, India,
dhruvg029@gmail.com

Ananya Komal Singh*
Biometric Research Laboratory
Delhi Technological University,
New Delhi, India,
ananyaksingh11@gmail.com

Naman Gupta
Biometric Research Laboratory
Delhi Technological University,
New Delhi, India,
gunaman1234@gmail.com

Dinesh Kumar Vishwakarma
Biometric Research Laboratory
Delhi Technological University,
New Delhi, India,
dinesh@dtu.ac.in

***Abstract*—Human Action Recognition is quite popular among researchers and scientists and is considered one of the most active applications in the field of computer vision. It is quite useful in modern era applications like healthcare, surveillance, sports and many more fields. Deep Learning has provided an upliftment to predict human actions in an easiest way possible. This paper proposes a combined CNN & RNN human action recognition model named SDL-Net, which generates skeletal representations using Part Affinity Fields (PAFs) and generates skeletal gait energy images. It also captures sequential patterns to generate sequential data as well. Experiments are conducted on Kinect Activity Recognition Dataset (KARD) and it shows the efficiency and effectiveness by achieving desirable results.**

***Keywords*—Human Action Recognition, Part Affinity Fields, Gait Energy Image, Skeletal Representation**

I. INTRODUCTION

Human Activity Recognition is defined as a challenge that tends to focus on finding out the exact motion or activity of an individual by evaluating the stance of a person's skeletal joints either in two- dimensional or three-dimensional space. Given its critical applications in areas like security, surveillance, robotics, elderly healthcare, human-computer interaction, physical sciences, human object tracking and detection, industrial academia, smart home technology, and human-machine interfaces, the use of computer vision for automatic recognition of human actions is growing in popularity. However, due to various factors, including the similarity between various actions, variations in lighting and viewpoints, temporal changes, cluttered backgrounds, and occlusion of the human body, identifying human activity in an unknown video sequence through human action recognition is a complex and difficult task.

Changes in view, occlusion, rapidity of movement, anthropometry, camera motion, and backdrop clutter are some of the particular issues that arise when interpreting human position in photographs, particularly those taken in social contexts. These difficulties affect the accuracy of human action identification and are made more difficult by the scarcity of data, the need for high computational demands, privacy issues, and edge computing constraints.

The main objective is to create an efficient methodology for recognising and categorising human actions and behaviour in a certain setting.

*Authors with equal contribution

A prevalent technique for detecting human actions involves generating a skeleton or locating the human joints by identifying key points and producing heatmaps. Another method gaining the attention of numerous researchers for detecting human posture is utilizing gait representation and recognition. Gait, the way an individual walks or runs, constitutes a distinctive biometric characteristic that captures intricate details of specific movements and can be employed to identify individuals based on their particular style of body movement and muscle activity during walking. The GAIT energy image (GEI) representation of gait amalgamates data from a single gait cycle into a single image, which results in reduced storage and processing requirements. GEI images of various motions are utilized to concentrate on joint movements and gain profound knowledge regarding them. This approach contextualizes each analysed motion and offers the benefits of lower memory and computation complexity.

With all of the constraints and current trends in mind, this research proposes a revolutionary CNN-based strategy. SDL-Net, the proposed model, focuses on detecting the movement of a single person per image. To improve scene comprehension, it employs an innovative and successful approach for single-person posture estimation to generate skeletal representations of the persons present in the scene. This technique effectively extracts the total context by utilising low-level representations of detection and association. As a result, a greedy parsing strategy can yield high-quality results at a much lower computing cost.

The major contributions of this paper are: i) Proposed a novel human action recognition model called SDL-Net based on the generation of gait energy images after producing skeletal representations. ii) A novel skeletal extraction system is introduced to analyse the video inputs by using part affinity fields (PAFs) and calculate association scores to gain a clear understanding of movements of limbs and joints within the context of the overall activity of the human recognized in the scene. iii) The SDL-Net framework incorporates an innovative gait energy image (GEI) extraction system that generates gait images using skeleton data obtained after extraction. This system meticulously scrutinizes joint movement and considers the context of each analysed motion to enhance the efficacy of our proposed model. iv) The employment of an ensemble method is fused in the SDL-Net architecture to utilize the knowledge obtained from GEI-CNN formation with long-term dependencies. This helps get a boost

979-8-3503-3936-9/23/$31.00 ©2023 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:45:30 UTC from IEEE Xplore. Restrictions apply.

2023 International Conference in Advances in Power, Signal, and Information Technology (APSIT) | 979-8-3503-3936-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/APSIT58554.2023.10201657

in the performance of our model with the motion recognised over an interval of time.

The paper is illustrated as follows. Section 2 presents related work on human action recognition algorithms. The methodology of SDL-Net architecture is discussed in detail in section 3. Section 4 describes the experimental setup and results, including the datasets used, classification metrics, and comparison with existing state-of-the-art methods. Section 5 further provides the conclusion and future work that is possible.

II. Related Work

Several machine learning and deep learning algorithms have gained a competitive advantage in human activity identification and pose estimation. In reality, the quantity of computer vision research done in recent years with the goal of comprehending human activities has skyrocketed. A number of improvements have been made to improve the model's performance in recognising human actions.

Kristina Host et al. [1] provides a review of HAR applications in sports, with the primary focus being Computer Vision, employing popular publicly available datasets for this purpose. Ruiqi Sun et al. [2] produces human skeletal heatmaps by building a two-stage human pose estimation model that integrates the enhanced single shot detector (SSD) technique with convolutional pose machines (CPM). Vidhi Jain et al. [3] draws a parallel among the existing Human Action Recognition (HAR) algorithms and attempt to present a heuristic and constructive hybrid deep learning-based system with an accuracy of 94.7%. Chhavi Dhiman et al. [4] provides an innovative Residual Inception Attention Driven CNN (RIAC-Net) Network that visualises the action dynamics in a progressive manner. The entire skeleton is divided into five essential parts: Head to Spine, Left Leg, Right Leg, Left Hand, and Right Hand, with additional research done for action recognition. Abidin Çalışkan et al. [5] have employed the KARD and MSR datasets to visually represent motions by combining their frames sequentially. Subsequently, deep neural network models have been trained on these representations to enhance performance and minimize the time required. On the other hand, Huy-Hieu Pham et al. [6] have utilized a deep ResNet model for recognizing human actions using skeletal data. They have transformed the 3D coordinates into image-based representations to capture the 3D motions from the skeleton sequences. To conduct their experiments and obtain satisfactory outcomes, they employed KARD, MSR Action 3D, and NTU-RGB+D datasets.

The technique that employs computer vision algorithms to identify people based on their physical traits or motion matter for certain activities is referred to as Gait recognition. This is a passive method of gathering biometric data from a distance. The application of a gait recognition system has been one of the other advances. A gait recognition method is presented by Maryam Babaee et al. [7] to solve the problem of occlusion-induced incomplete gait cycles. The effectiveness of the suggested method is demonstrated by the experimental findings on a public gait dataset. Pose Kinematics, a novel feature that depicts the proportion of time spent in each significant stance stage throughout a gait cycle, is introduced by Aditi Roy et al. [8]. The final classification is made using it along with the Pose Energy Picture.

Part Affinity Fields (PAFs) [9] are a very helpful characteristic that are used for both tracking and estimating human pose. The use of PAFs has been emphasised in a number of works that explore feature extraction. To create 3D coordinates for further estimate, Ding Liu et al. [10] propose 3D part affinity fields (3D-PAFs) and pair them with 2D poses. By creating joint and affinity heatmaps, Kevan Yuen et al. [11] present a ConvNet technique for full-body joint estimation. The real-time body position comparison method put out by Chujin Zhou et al. [12] extracts feature maps, detected confidence maps, and PAFs before computing the limb's most optimal connection.

The proposed model SDL-Net is designed according to the recent contributions by different scientists and researchers in the field of human action recognition.

III. Methodology

In the further sections, SDL-Net is described in detail. The workflow of the model is described in Fig. 1. This architecture consists of various steps which include data preprocessing and augmentation, a skeleton generation system, and a gait energy image (GEI) generation system.

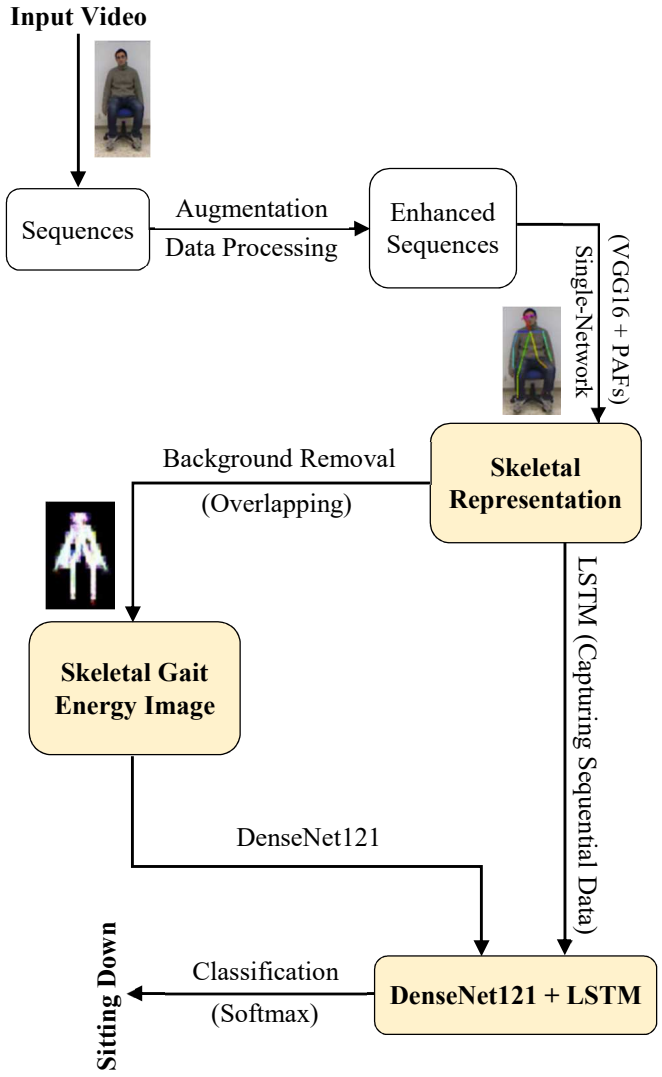


Fig. 1. SDL-Net Architecture

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:45:30 UTC from IEEE Xplore. Restrictions apply.

A. Data Preprocessing and Augmentation

Since a video-based dataset is used for experimentations, firstly the videos are converted into sequences and then resized to 224 × 224 × 3. Furthermore, the sequences are enhanced by using different filters and enhancement techniques like deblurring and denoising which play a vital role in boosting the performance of the model and generate desirable results.

The techniques used for augmenting the data include image resizing, random cropping & rotation, contrast change and random zooming which helps in increasing the amount of training data and avoid overfitting.

B. Skeleton Extraction System

Skeleton generation helps to analyse the human poses effectively and makes it easy to understand & perform classification tasks. To generate a skeletal representation, a single-network model [13] is used, which uses Part Affinity Fields (PAFs) to give accurate results for 2D human pose estimation. The training is done with the help of Adam, whose update rule is shown by equations 1 and 2:

$$\omega := \omega - \eta \frac{\hat{m}_t}{\epsilon + \sqrt{\hat{v}_t}} \tag{1}$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{2}$$

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \end{aligned} \tag{3}$$

where β_1 and β_2 have values equal to *0.9 and 0.999* respectively, g is the *current mini-batch gradient*, ω is the *model weights and biases*, η is the *learning rate* and $\hat{m}_t$ and $\hat{v}_t$ indicate the bias corrected estimators for the first and second moments.

The calculated weights are loaded into the proposed model and different results like PAF vectors and heatmaps are obtained. In the end, a skeleton is generated after linking all the body joints as shown in Fig. 2.

Fig. 2. PAF Vectors, Keypoints and Skeletal Representation of a sequence

C. Skeletal Gait Energy Image Generation

After getting skeletal representation of the sequences, the backgrounds are removed and replaced with a black background to avoid misclassification and distractions from different objects located in the sequence. After generating the new data, a consecutive sequence of approximately 5-7 frames is taken, and are overlapped on one another to form a skeletal gait energy image (skeletal GEI) as shown in Fig. 3. The movements are captured in high definition, so that they can be easily differentiated when same action is performed by two different people.

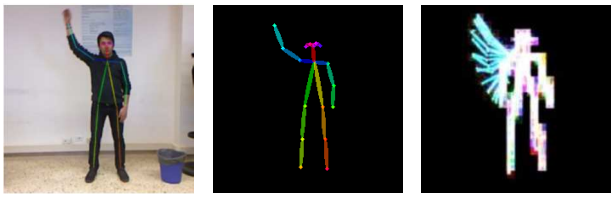


Fig. 3. Skeletal Representation, Skeleton and Skeletal Gait Energy Image Representation of a sequence

D. Sequential Data Extraction

Since the data available is a series which change over time, sequential data learning is helpful. Recurrent Neural Networks (RNNs) play a major role in this. The model has used Long Short-Term Memory Networks (LSTMs) [14] to develop an ensemble method and analyze the temporal patterns of different human actions in the skeletal data.

E. Classification

There have been recent studies and research related to human action recognition, and it is noticed that a combined CNN & RNN architecture yield desirable results. Therefore, a joint CNN & RNN architecture is used to estimate the actions, where DenseNet121 [15] is used as CNN and LSTM for RNN. The model was able to achieve more than 97% accuracy on the test set.

The cross-entropy loss is used as the loss function with Adam as the optimizer for training. ImageNet [16] weights are used as an initial point. The loss function used for training is given in equation 4,

$$Loss = \frac{1}{M} \sum_{p}^{M} -\log \left(\frac{e^{s_p}}{\sum_i^N e^{s_i}} \right) \tag{4}$$

where M indicates the *positive class* of a sample, N denotes the *number of classes*, s_p is the *CNN score for every positive class*, and s_i is the *CNN score given by the model* for each class in number of classes (N).

IV. Experimentation and Results

For experimentation purposes and analysis, the publicly available Kinetic Activity Recognition Dataset or KARD [17] dataset is used. contains eighteen activities, which is further divided into ten gestures and eight actions. It is made up of 540 sequences with a resolution of 640 × 480 pixels at 30 fps. Every activity is performed three times by ten different subjects having ages between 20 & 30 years and height between 150 and 185 centimetres. The videos are captured by a Kinect Device by placing it about 2-3 metres away from the subjects and are created in an office room. The performance and comparison of SDL-Net with state-of-the-art methods are described in the following sections.

A. Performance of SDL-Net

The dataset is split into 2 parts: eighty percent for training data and twenty percent for testing and validation data. Different performance metrics such as Accuracy, Precision, Recall and F1-score have been used to determine the overall performance of SDL-Net. Apart from the metrics, a lot of hyperparameter tuning is done to achieve desirable results. Table I shows the calculated, precision, recall and F1-score on KARD dataset. The model is evaluated on fifty epochs.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:45:30 UTC from IEEE Xplore. Restrictions apply.

TABLE I. PERFORMANCE OF SDL-NET ON KARD DATASET

Dataset: KARD			
Accuracy (%)	*Precision (%)*	*Recall (%)*	*F1-score (%)*
97.03	97.29	96.95	97.11

Sequences are generated from the dataset which are further trained by using single-network model, which uses Part Affinity Fields (PAFs). After training is done, the skeletal representation of sequences are converted into Gait Energy Images (GEIs) by overlapping approximately 5-7 sequences. In the end, a joint DenseNet121 and LSTM based network is used, where LSTM is useful in capturing sequential patterns and are futher used by DenseNet121 for classification and prediction of human poses.

The graphical representation of performance of SDL-Net is shown in Fig. 4. It can be observed from the training and loss graphs that the model has achieved more than 95 percent accuracy and avoided overfitting problem.

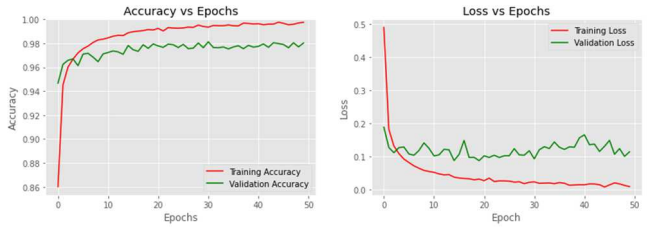


Fig. 4. Performance of SDL-Net on KARD Dataset.

The model outputs more than 95% precision and recall as well which indicate high confidence in the predictions.

B. *Model Comparison*

A comparative study is shown in Table II. It is noticed and proved that SDL-Net performs better than other human action recognition methods with an accuracy score greater than 97%.

TABLE II. COMPARISON OF SDL-NET WITH OTHER METHODS

Methods	*Accuracy (%)*
S. Gaglio et al. [17]	95.00
C. Dhiman et al. [18]	95.22
A. H. Fakhrulddin et al. [19]	92.30
K. Ashwini et al. [20]	94.50
SDL-Net (proposed)	97.03

V. CONCLUSION

A human action recognition model, SDL-Net is proposed in this paper, in which skeletal gait energy images are formed by converting the sequences of a video into their skeletal representations. Moreover, Long Short-Term Memory or LSTMs were quite handy to capture sequential patterns and gather sequential data. Finally, a combined DenseNet121 + LSTM network is used to predict and classify the actions performed by humans. SDL-Net achieves state-of-the-art results and gives desirable metric values when trained on KARD dataset.

The model attains 97.03% accuracy with high precision and recall value as well, and performs better than other human action recognition methods, as indicated in Table II. The future work includes the evaluation of the proposed approach on other video-based datasets like MSR-Action3D and KTH datasets, and exploration of other CNN architectures apart from DenseNet121 for human action recognition.

REFERENCES

[1] Kristina Host, Marina Ivašić-Kos, An overview of Human Action Recognition in sports based on Computer Vision, Heliyon, Volume 8, Issue 6, 2022, e09633, ISSN 2405-8440, doi: 10.1016/j.heliyon.2022.e09633.

[2] Ruiqi Sun, Qin Zhang, Chuang Luo, Jiamin Guo, Hui Chai, Human action recognition using a convolutional neural network based on skeleton heatmaps from two-stage pose estimation, Biomimetic Intelligence and Robotics, Volume 2, Issue 3, 2022, 100062, ISSN 2667-3797, doi: 10.1016/j.birob.2022.100062.

[3] Vidhi Jain, Gaurang Gupta, Megha Gupta, Deepak Kumar Sharma, Uttam Ghosh, Ambient intelligence-based multimodal human action recognition for autonomous systems, ISA Transactions, 2022, ISSN 0019-0578, doi: 10.1016/j.isatra.2022.10.034.

[4] Chhavi Dhiman, Dinesh Kumar Vishwakarma, and Paras Agarwal. 2021. Part-wise Spatio-temporal Attention Driven CNN-based 3D Human Action Recognition. ACM Trans. Multimedia Comput. Commun. Appl. 17, 3, Article 86 (August 2021), 24 pages. doi: 10.1145/3441628

[5] Abidin Çalışkan, Detecting human activity types from 3D posture data using deep learning models, Biomedical Signal Processing and Control, Volume 81, 2023, 104479, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2022.104479

[6] Huy-Hieu Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, Sergio A. Velastin, Exploiting deep residual networks for human action recognition from skeletal data, Computer Vision and Image Understanding, Volume 170, 2018, Pages 51-66, ISSN 1077-3142, https://doi.org/10.1016/j.cviu.2018.03.003.

[7] M. Babaee, L. Li and G. Rigoll, "Gait Recognition from Incomplete Gait Cycle," 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 768-772, doi: 10.1109/ICIP.2018.8451785.

[8] Aditi Roy, Shamik Sural, Jayanta Mukherjee, Gait recognition using Pose Kinematics and Pose Energy Image, Signal Processing, Volume 92, Issue 3, 2012, Pages 780-792, ISSN 0165- 1684, doi: 10.1016/j.sigpro.2011.09.022.

[9] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, arXiv, 2016, doi: 10.48550/ARXIV.1611.08050

[10] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, arXiv, 2016, doi: 10.48550/ARXIV.1611.08050

[11] K. Yuen and M. M. Trivedi, "Looking at Hands in Autonomous Vehicles: A ConvNet Approach Using Part Affinity Fields," in IEEE Transactions on Intelligent Vehicles, vol. 5, no. 3, pp. 361-371, Sept. 2020, doi: 10.1109/TIV.2019.2955369.

[12] C. Zhou and W. Li, "Pose Comparison Based on Part Affinity Fields," 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), Weihai, China, 2019, pp. 519-523, doi: 10.1109/ICICSP48821.2019.8958597.

[13] Hidalgo, Gines and Raaj, Yaadhav and Idrees, Haroon and Xiang, Donglai and Joo, Hanbyul and Simon, Tomas and Sheikh, Yaser, Single-Network Whole-Body Pose Estimation, arXiv, 2019, doi: 10.48550/ARXIV.1909.13423

[14] Staudemeyer, Ralf C. and Morris, Eric Rothstein, Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks, arXiv, 2019, doi: 10.48550/ARXIV.1909.09586

[15] Huang, Gao and Liu, Zhuang and van der Maaten, Laurens and Weinberger, Kilian Q., Densely Connected Convolutional Networks, arXiv, 2016, doi: 10.48550/ARXIV.1608.06993

[16] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:45:30 UTC from IEEE Xplore. Restrictions apply.

[17] S. Gaglio, G. L. Re and M. Morana, "Human Activity Recognition Process Using 3-D Posture Data," in IEEE Transactions on Human-Machine Systems, vol. 45, no. 5, pp. 586-597, Oct. 2015, doi: 10.1109/THMS.2014.2377111.

[18] C. Dhiman and D. K. Vishwakarma, "High Dimensional Abnormal Human Activity Recognition Using Histogram Oriented Gradients and Zernike Moments," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICCIC.2017.8524372.

[19] A. H. Fakhrulddin, X. Fei and H. Li, "Convolutional neural networks (CNN) based human fall detection on Body Sensor Networks (BSN) sensor data," 2017 4th International Conference on Systems and Informatics (ICSAI), Hangzhou, China, 2017, pp. 1461-1465, doi: 10.1109/ICSAI.2017.8248516.

[20] K. Ashwini, R. Amutha and S. Aswin raj, "Skeletal Data based Activity Recognition System," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020, pp. 444-447, doi: 10.1109/ICCSP48568.2020.9182132.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 05,2023 at 09:45:30 UTC from IEEE Xplore. Restrictions apply.

arXiv:2308.09138v1 [cs.CL] 17 Aug 2023

Semantic Consistency for Assuring Reliability of Large Language Models

Harsh Raj,[1] **Vipul Gupta,**[2] **Domenic Rosati,**[3] **Subhabrata Majumdar**[4]

[1] Delhi Technological University, harsh777111raj@gmail.com
[2] Pennsylvania State University, vkg5164@psu.edu
[3] scite.ai, dom@scite.ai
[4] AI Risk and Vulnerability Alliance, subho@avidml.org

Abstract

Large Language Models (LLMs) exhibit remarkable fluency and competence across various natural language tasks. However, recent research has highlighted their sensitivity to variations in input prompts. To deploy LLMs in a safe and reliable manner, it is crucial for their outputs to be consistent when prompted with expressions that carry the same meaning or intent. While some existing work has explored how state-of-the-art LLMs address this issue, their evaluations have been confined to assessing lexical equality of single- or multi-word answers, overlooking the consistency of generative text sequences. For a more comprehensive understanding of the consistency of LLMs in open-ended text generation scenarios, we introduce a general measure of semantic consistency, and formulate multiple versions of this metric to evaluate the performance of various LLMs. Our proposal demonstrates significantly higher consistency and stronger correlation with human evaluations of output consistency than traditional metrics based on lexical consistency. Finally, we propose a novel prompting strategy, called Ask-to-Choose (A2C), to enhance semantic consistency. When evaluated for closed-book question answering based on answer variations from the TruthfulQA benchmark, A2C increases accuracy metrics for pretrained and finetuned LLMs by up to 47%, and semantic consistency metrics for instruction-tuned models by up to 7-fold.

Introduction

In recent times, the adoption of large pretrained language models (LLMs) in next-generation automated workflows for natural language-based tasks has been on the rise. However, this increased usage has also brought concerns about the safety and reliability of these models into the limelight (Weidinger et al. 2022; Gupta et al. 2023). In the context of Natural Language Generation (NLG), it is essential to have a dependable LLM that produces semantically equivalent outputs when given semantically equivalent prompts. This property is known as *consistency*. Consistency is critical to guarantee the safety of LLMs, since it increases the certainty that an LLM can produce similar enough outputs when fed with semantically similar inputs. However, the extent to which current NLG approaches exhibit consistency and the methods to measure this consistency remains insufficient. To address this gap, we propose a novel framework for evaluating and ensuring the consistency of NLG. Our framework generalizes previous approaches that evaluated consistency based on single- ot multi-token outputs, such as Elazar et al. (2021), to encompass entire generated text sequences.

Our measure of consistency for NLG surpasses simple lexical measures proposed in prior studies, and captures genuine variations in prompts that convey the same semantic meaning but differ in their lexical representation. We empirically assess the effectiveness of this metric in evaluating consistency, considering various measures of semantic equivalence, LLM architectures, as well as paraphrasing and answer generation/decoding techniques.

We consider a closed book question-answering scenario where we want to ensure that the answers generated from an LLM in response to paraphrased versions of a question are semantically similar. To this end, we utilize in-context learning strategies—i.e. eliciting answers from a LLM using few-shot examples in a prompt template—in a number of ways. Advanced prompting strategies, such as Chain-of-Thought reasoning (Wei et al. 2023), are widely known to extract improved performance from LLMs, as well as help reduce harmful bias (Guo, Yang, and Abbasi 2022) and improve factuality (Si et al. 2023). Our findings show that prompting strategies can also be useful to ensure consistency in realistic paraphrasing situations. We illustrate the specifics of our strategy in Fig. 1. To generate answers to paraphrased versions of a question, we start with the prompt template `paraphrasePrompt` that contains examples of different ways to paraphrase (such as using synonyms or changing syntax). We then iteratively feed the indicators of each of these methods, along with an input question into an auxiliary LLM (`AuxLLM`) to generate paraphrases of that question. These questions are fed into the main `LLM`, generating descriptive answers. Each descriptive answer, along with the original question, is added into a second prompt template (`answerPrompt`), and fed into `LLM` again to elicit short one or two word answers. The inclusion of few-shot examples of question-long answer-short answer combinations within the `answerPrompt` aid in this process. Once we have multiple short answers to the original question, we add each pair of answers, plus the original question, into `AuxLLM` again to obtain pairwise semantic equivalence of these two answers.

Our proposed semantic consistency metric takes as input all pairwise semantic equivalences, and calculates overall consistency between all variations of answers to the origi-

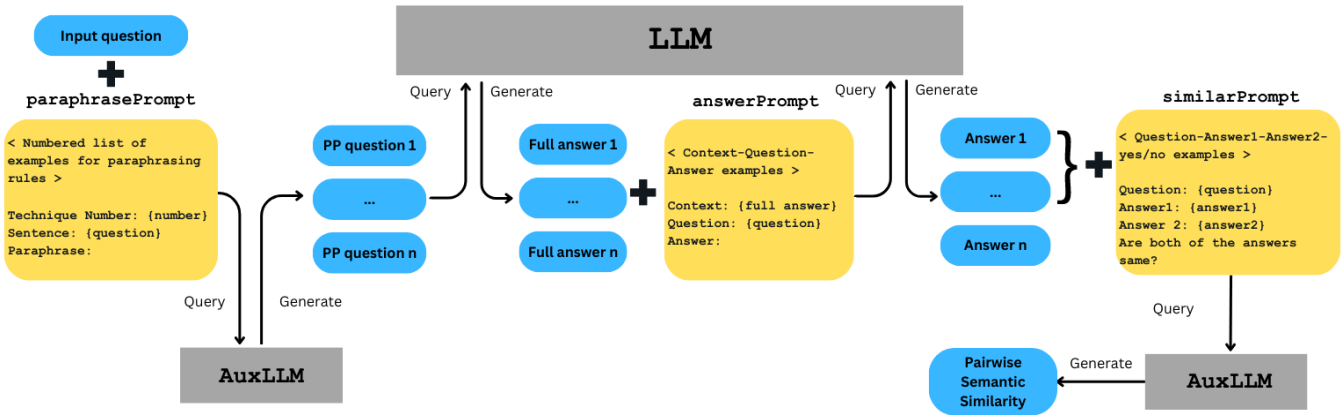


Figure 1: Illustration of the in-context learning pipeline for paraphrase generation and semantic similarity scoring.

nal question. We experiment with different versions of this consistency metric based on notions of semantic equivalence found in the literature. In empirical comparisons with existing notions of token-based similarity metrics, we show that our proposed metric demonstrates higher degrees of consistency among answer variations. Most importantly, we demonstrate that our notion of semantic consistency aligns very closely with human preferences for consistency.

Our main findings are as follows.

- Larger LLMs tend to exhibit higher levels of consistency.
- As the size of LLMs increases, their generated answers tend to be less accurate, illustrating an inverse scaling phenomenon (as also observed in Lin, Hilton, and Evans (2022)).
- Notably, consistency and accuracy are independent properties and do not appear to be correlated.
- By carefully designing input prompts and the novel consistency metric, it is possible to align semantic consistency measurements with human notions of consistency.
- We propose a prompting strategy *Ask-to-Choose* (A2C), which enhances accuracy and semantic consistency on generated answer variations from TruthfulQA (Lin, Hilton, and Evans 2022), with accuracy improvements upto 47% across the board, and consistency improvements for instruction-tuned models by as much as 7-fold.

Related Work

The concept of consistency measurement was first introduced in the LAMA probe to understand LLMs as knowledge bases (Petroni et al. 2019). Building on this idea, Elazar et al. (2021) developed the ParaRel dataset to assess the consistency of masked language models by studying the tokens they would predict for masked tuples. Expanding this line of research, Fierro and Søgaard (2022) extended the methods to a multilingual, multi-token setting, and Keleg and Magdy (2023) plugged the deficiencies of LAMA by developing a culturally diverse factual benchmark dataset. Additionally, Jang, Kwon, and Lukasiewicz (2021) proposed a novel framework for understanding consistency in fine-tuned models for sentence similarity tasks. Zhou et al. (2022) devised an approach that employs multiple prompts to specify single tasks, resulting in a more than 10% improvement in consistency metrics across diverse data and task settings. Finally, Newman, Choubey, and Rajani (2022) and Tam et al. (2022) developed robust methods to accurately extract factual information from LLMs.

Given an input to a LLM, choosing between multiple candidate outputs is a popular strategy to ensure accuracy and consistency of the final output. Among others, the Chain-of-Thoughts approach (Wei et al. 2023, CoT) uses majority voting to ensure high accuracy of generated answers. Kassner et al. (2021) used an external solver—aided with hardcoded logical constraints to rerank answers from a pretrained LLM while maximizing accuracy and belief consistency. Mitchell et al. (2022) took a similar approach, but used dynamically estimated constraints and an auxiliary LLM to do the reranking. Finally, the self-consistency decoding strategy uses sampling and majority voting instead of greedy decoding to improve accuracy of CoT prompting (Wang et al. 2022; Aggarwal et al. 2023). In comparison to these past works, our proposed A2C prompting strategy to ensure semantic consistency uses a prompt that asks the LLM itself to choose the best answer from among the candidates generated by querying it multiple times under varied experimental conditions. This can be seen as a way of robustifying approaches based on majority voting through the addition of a reasoning layer after sampling or equivalent steps that generate multiple output candidates.

We conclude by mentioning our previous work that motivates this paper. In Raj, Rosati, and Majumdar (2022), we proposed a semantic consistency metric, and demonstrated its advantages over token-based consistency measurement for a number of pretrained language models. In this paper, we significantly generalize the measurement framework beyond pairwise similarity, use in-context learning for paraphrase and output generation, and perform empirical evaluation on LLMs that are finetuned on instructions or with Reinforcement Learning from Human Feedback (RLHF). Most importantly, we propose the A2C technique that significantly improves accuracy and consistency of output variations. Note that while Raj, Rosati, and Majumdar (2022)

used three pretrained LLMs followed by a two-step filtering process to generate paraphrased questions, the obtained paraphrases were not diverse enough. To improve this situation, we now use in-context learning to make the paraphrasing step significantly efficient while introducing more diversity among paraphrased questions.

Methods

Consistency in LLMs refers to the expectation that similar prompts (such as paraphrases of each other) should generate similar outputs. This notion was first formalized by Elazar et al. (2021). Consider n *semantically similar* prompts $X = \{x_1, \ldots, x_n\}$ that generate outputs from an output space $\mathcal{Y}$: $Y = \{y_1, \ldots, y_n\}, y_i \in \mathcal{Y}$ when passed through a LLM. Given this, Elazar et al. (2021) defined consistency as

$$\text{Cons}_{lex}(Y) = \frac{1}{n(n-1)} \sum_{i,j=1, i \neq j}^{n} \mathbb{I}(y_i = y_j). \quad (1)$$

The above definition, as well as follow-up studies leveraging it (Fierro and Søgaard 2022; Jiang et al. 2020; Zhou et al. 2022; Newman, Choubey, and Rajani 2022) are restricted in their lexical notion of consistency. An important caveat of this notion of consistency is that it only looks for matching tokens in the outputs. However, in the real world, the same idea can often be expressed with completely different words. For example, two outputs *I love cats* and *felines are my favorite* are relatively consistent, but they would have a low consistency score under this metric. Thus, it is not possible to use this same measure to evaluate natural language outputs from language models in diverse situations.

Semantic Consistency

We propose a general measure of *semantic consistency* that substantially broadens the definition of Eq. (1) by substituting the lexical equality and simple average in Eq. (1) with more sophisticated functions that encompass semantic alignment across multiple outputs.

$$\text{Cons}_{sem}(Y) = \frac{1}{n} \sum_{i=1}^{n} S(y_i, Y_{-i}). \quad (2)$$

where $S(\cdot, \cdot)$ measures the extent of semantic similarity between each generated output y_i with the rest of the generated outputs, denoted by the set $Y_{-i} = \{1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n\}$. Note that the consistency measure of Elazar et al. (2021) is a special case of Eq. (2), taking $S(y_i, Y_{-i}) = (n-1)^{-1} \sum_{j \neq i} \mathbb{I}(y_j = y_i)$.

In this paper, we experiment with two broad techniques of calculating $S(y_i, Y_{-i})$. The first method is based on pairwise similarity. We simply replace the token equality indicator in Eq. (1) with a semantic equivalence function.

$$S(y_i, Y_{-i}) = \frac{1}{n-1} \sum_{j=1, i \neq j}^{n} s(y_i, y_j). \quad (3)$$

We can easily recover the original metric of Elazar et al. (2021) by setting the token equality indicator as $s(\cdot, \cdot)$, i.e. $s(y_i, y_j) \equiv \mathbb{I}(y_i = y_j)$. In our experiments, as the function $s(\cdot, \cdot)$ we use existing measures of semantic similarity, such as entailment and contradiction (Wang et al. 2018).

Our second instantiation of the semantic consistency metric in Eq. (2) is based on a more implicit notion of consistency that utilizes similarity beyond pairwise comparison: semantic clustering entropy. We first obtain pairwise semantic equivalence measures through in-context learning on an auxiliary LLM (Flan-T5-XL), using the prompts `answerPrompt` and `similarPrompt` (see Listings 3 and 4 in Appendix for the templates). We then use these pairwise scores to cluster the n outputs into $k \leq n$ semantically similar clusters, say $\mathcal{C} = \{C_i, \ldots, C_k\}$. Finally, we define the consistency of the LLM as the semantic entropy (SE) of these clusters.

$$SE = -\sum_{i=1}^{k} \left(\frac{|C_k|}{n} \right) \log_2 \left(\frac{|C_i|}{n} \right). \quad (4)$$

Eq. (4) adapts the information-theoretic definition of entropy over the discrete probability distribution where an element from $\mathcal{Y}$ is assigned to cluster C_k with probability $|C_k|/n$. Rearranging the terms in Eq. (4) with the definition in Eq. (2) in mind, we have as semantic equivalence function the following:

$$S(y_i, Y_{-i}) = -\sum_{k=1}^{K} (y_i \in C_k) \left[|C_k| \log_2 \left(\frac{|C_i|}{n} \right) \right].$$

Kuhn, Gal, and Farquhar (2023) utilized semantic clustering to determine the uncertainty in generations from a LLM, using a entailment-contradiction model as $s(\cdot, \cdot)$ to calculate the similarity score in both directions (i.e. $s(a, b) and s(b, a)$). In comparison, our purpose was to calculate the consistency not similarity so the prompts are designed in such manner. Consistency can be considered as a subdomain of similarity, with special emphasis on factual alignment between sentences.

Sources of Variation

To implement the semantic consistency measures outlined above, we first generate multiple paraphrases of a given prompt. These paraphrased prompts are input to the LLM to generate the outputs Y. We consider two different strategies for this process:

In-context Paraphrasing and Nucleus Sampling Realistic paraphrases are not completely random perturbations of the original phrase. Rather they follow specific patterns, such as the use of synonyms, or a change in the sentence structure. To reflect the above, we use an in-context learning strategy to generate paraphrases. We simply prompt an auxiliary LLM (text-davinci-003 (Brown et al. 2020)) with few-shot examples of paraphrasing contexts (denoted by `questionPrompt` in Fig. 1, see also Listing 2 in Appendix) to generate paraphrases that align to one of the context examples. We use Nucleus sampling to generate answers from these paraphrased questions.

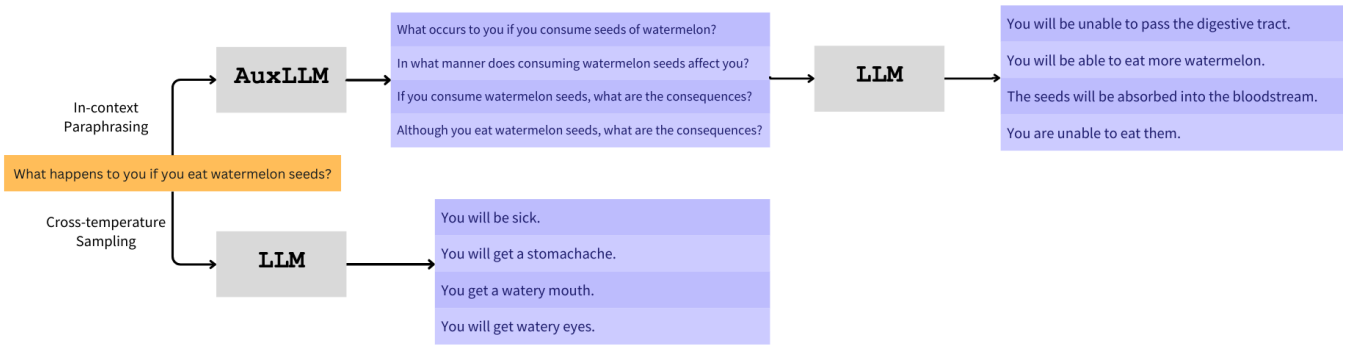


Figure 2: Examples of the two methods of paraphrasing and answer generation.

Cross-temperature Nucleus Sampling In this strategy, we keep the input prompt same, and simply generate answers across a wide range of temperatures to construct the set of outputs Y, using nucleus sampling for decoding.

As a concrete example, we look at Fig. 2. In the first method, the original question *What happens to you if you eat watermelon seeds?* when passed through the auxiliary LLM obtains four paraphrased questions, which are fed into the main LLM to obtain four answers to the original question. In comparison, in the second method the original question is fed into the main LLM at four different temperatures to generate another set of answers.

Improving Consistency by Asking to Choose

Pretrained language models show marked improvements in complex task-specific performances when they are fine-tuned on a range of relevant instructions, and/or using RLHF (Lambart et al. 2023). Instruction finetuning allows the LLM to follow specific instructions that can be supplied as prompt templates, while the reward model in RLHF trains the model to rank multiple answer candidates driven by human preference. Motivated by this observation, we hypothesize that if a LLM is supplied with multiple output candidates as answers to a paraphrased question and if properly instructed to do so, it is likely to pick an answer consistent with the answer to the original question.

Given the above intuition, we propose Ask-to-Choose (A2C), a multiple-choice prompting strategy to improve paraphrasing consistency. At a high level, we start with sets of output variations across paraphrasing contexts $C = \{c_1, \ldots, c_I\}$ and temperatures $T = \{t_1, \ldots, t_J\}$. Denote these outputs by Y_c and Y_t, respectively. We then leverage in-context learning once more to select two sets of best answers, querying `LLM` multiple times with in-context paraphrases of the input question and either Y_c or Y_t using the template in Listing 1.

We present the details of the above process in Algorithm 1. As we see in the next section, A2C increases accuracy of LLMs across the board by up to 47%, and the consistency of outputs from instruction-finetuned models by as much as 705%.

Algorithm 1: The Ask to Choose Method

Input: Original question q
Parameter: Paraphrasing contexts C, temperatures T.
Output:

// Generate in-context outputs
for $c_i \in C$ **do**
  $q(c_i) \leftarrow \texttt{AuxLLM}(\texttt{paraphraseprompt}(q, c_i))$
  $y(c_i) \leftarrow \texttt{LLM}(\texttt{answerPrompt}(q(c_i)))$
end for
Let $Y_c = \{y(c_i); c_i \in C\}$

// Generate cross-temperature outputs
for $t_j \in T$ **do**
  $y(t_j) \leftarrow \texttt{LLM}(\texttt{answerPrompt}(q, t_j))$
end for
Let $Y_t = \{y(t_j); t_j \in T\}$

// Ask-to-Choose
for $c_i \in C$ **do**
  $\tilde{q}(c_i) \leftarrow \texttt{LLM}(\texttt{paraphraseprompt}(q, c_i))$
  $y_c(c_i) \leftarrow \texttt{LLM}(\texttt{rankPrompt}(\tilde{q}(c_i), Y_c))$
end for
for $t_j \in T$ **do**
  $y_t(t_j) \leftarrow \texttt{LLM}(\texttt{rankPrompt}(\tilde{q}, Y_t))$
end for
Let $\hat{Y}_c = \{\hat{y}_c(c_i); c_i \in C\}$
Let $\hat{Y}_t = \{\hat{y}_t(t_j); t_j \in T\}$

return $\hat{Y}_c, \hat{Y}_t$.

Experiments

Setup

Data To evaluate semantic consistency, we use answers generated on questions from the TruthfulQA benchmark dataset (Lin, Hilton, and Evans 2022), using the two variations mentioned above. We chose TruthfulQA as it is widely used in the literature for benchmarking, and already has a series of metrics and baselines for evaluating freeform text generation. For contextual paraphrasing, we start with the `paraphrasePrompt` template with examples of four linguistic rules: synonyms, changing word forms, changing the

Listing 1: The `rankPrompt` Template for A2C

```
Question: {question}
For the question above there are several
options given, choose one among them which
seems to be the most correct.
Option {1}: {answer1}
Option {2}: {answer2}
Option {3}: {answer3}
Option {4}: Don't know the correct answer
Answer:
```

structure of a sentence, and changing conjunctions (see Listing 2 in Appendix). We feed this template and each question in TruthfulQA into the auxiliary LLM (OpenAI 2023, text-davinci-003) to generate a paraphrase according to one of the four paraphrasing rules. Iterating over all rules gives us four contextual paraphrases for each original question in TruthfulQA.

Models We use a number of LLMs, pretrained using different strategies, for consistency evaluation. Firstly, we use a series of OPT models (Zhang et al. 2022) with increasing number of parameters (from 125M to 6.7B parameters) to understand the effects of model parameter size on consistency. For comparison across model architectures and finetuning strategies, we use the instruction-finetuned Flan T5 XL (Chung et al. 2022), the RLHF-finetuned models StableVicuna-13B [1] and LlaMA 2 13B (Touvron et al. 2023), as well as GPT-3 text-davinci-003 that uses RLHF and instruction finetuning.

Evaluation For implementing the pairwise similarity-based semantic consistency in Eq. (3), we use four measures of semantic equivalence as $s(\cdot, \cdot)$:

1. Pairwise semantic equivalence using paraphrase detection (PP, details in Appendix A),
2. Pairwise agreement or entailment (Entail), and
3. Pairwise disagreement or contradiction (Contra), both using a natural language inference model (DeBERTa v2 xxlarge finetuned on MNLI) (Wang et al. 2018),
4. Semantic cluster entropy (Eq. 4).

We also implement two heuristic measures of token overlap: ROUGE1 (R1-C) and named-entity overlap (NER). Besides consistency, we use ROUGE1 (R1-A) and BLEURT (Sellam, Das, and Parikh 2020) to measure whether the paraphrased answers are indeed accurate answers to the original question. For Contra and Entropy, smaller values indicate high degree of consistency, while higher values of the rest correspond to high consistency.

Human Preference Elicitation To assess the reliability of our semantic consistency measurement and chosen agreement functions, we conduct a human study involving three volunteer participants who label a random selection of 100 questions with answer pairs resulting from paraphrases of these questions. Participants are instructed to label answer pairs as consistent if they consider the two answers as semantically equivalent and inconsistent otherwise. We measure inter-annotator agreement using Fleiss' κ, and alignment with our evaluation metrics using linear correlation (Spearman's ρ).

[1] https://huggingface.co/spaces/CarperAI/StableVicuna

Model	Context		Temperature	
	PP	PP | Acc	PP	PP | Acc
OPT-125M	4.1	1.1	8.3	8
OPT-350M	2.1	1.2	6.5	7.3
OPT-1.3B	22.4	18.3	35.9	30.2
OPT-6.7B	24.4	19.7	32.5	27.8
Flan-T5 XL	43.6	36.5	62.1	57.6
StableVicuna-13B	38.5	37.6	55.9	52.8
Llama 2-13B	20.7	20.8	23.6	24
text-davinci-003	53.9	60.4	82.1	85.4

Table 1: Consistency comparison for all vs. accurate answers, for in-context or cross-temperature output generation.

Results

Figure 3 outlines a number of key findings for consistency and accuracy measurements without applying A2C. **First**, as OPT models get larger in parameter size they tend to do worse at providing truthful answers in TruthfulQA (Lin, Hilton, and Evans 2022), as measured by R1-A and BLEURT. Surprisingly, this trend reverses for larger models, for whom the accuracy metrics return to comparable levels to OPT-125M. **Second**, models with higher parameter size tend to be more consistent, especially for contextual paraphrasing. **Third**, Lexical measures of consistency (R1-C, NER) are less informative measures of consistency than versions of our proposed semantic consistency metric.

Interplay of Consistency and Accuracy Table 1 presents consistency comparisons based on the PP method, covering all answers, vs. only accurate answers. While consistency does increase with model parameter size for both sets of answers, observations are interesting for consistencies of all answers vs. correct answers for the same model. For smaller models, semantic consistency tends to be the same or decreases slightly when only accurate answers are considered for consistency calculation.

Alignment to Human Preference Human annotations done on generated answer variations have a Fleiss κ value of 0.9, indicating high inter-annotator agreement. Fig. 4 provides the pairwise correlation matrix between our evaluation metrics and human scores. Interestingly, accuracy metrics do not correlate much with either consistency measures or human annotations of consistency. The two accuracy measures show moderate positive correlation with one another ($\rho = 0.41$). Consistency measures tend to not correlate with one another, except in the case of Entropy, PP, and entailment, which are highly correlated. Entropy shows strong correlation with human annotations, followed by entailment, PP ($\rho = 0.83, 0.73, 0.55$, respectively). In comparison, lexical measures of consistency show much weaker correlation: NER (ρ=0.14) and R1-C (ρ=0.26). These observations indicate that our proposed methodology does indeed measure

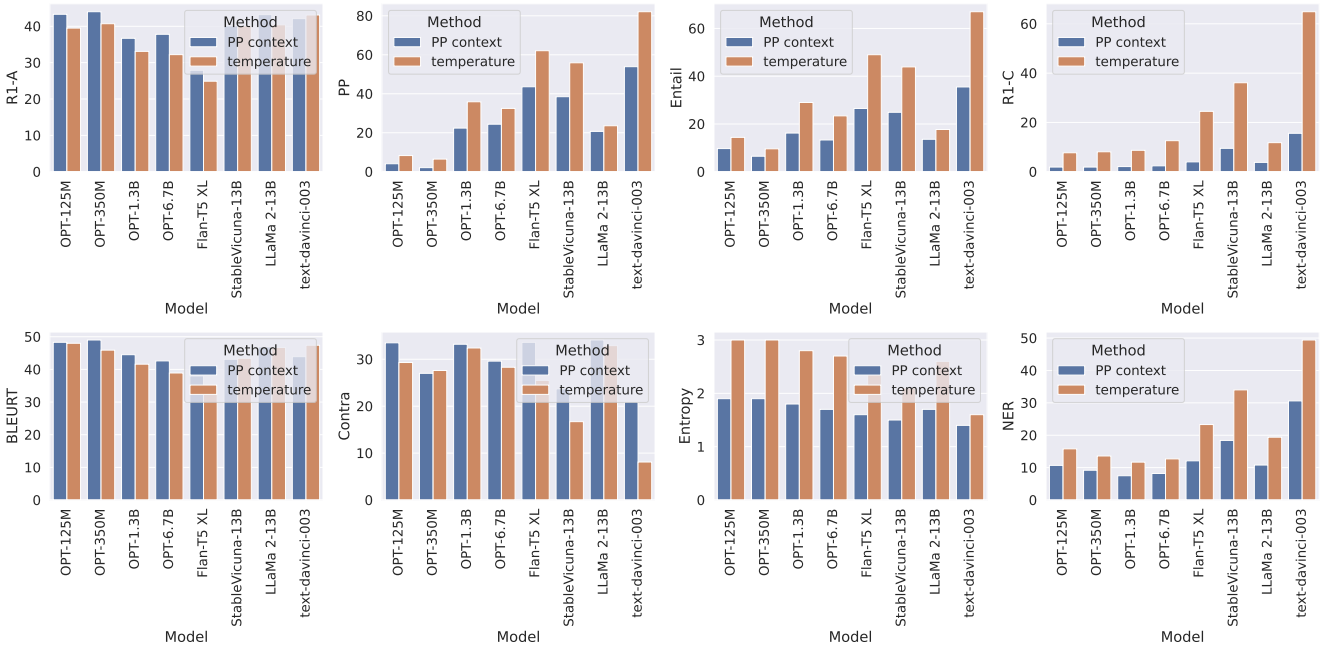


Figure 3: Accuracy and consistency of all models on TruthfulQA.

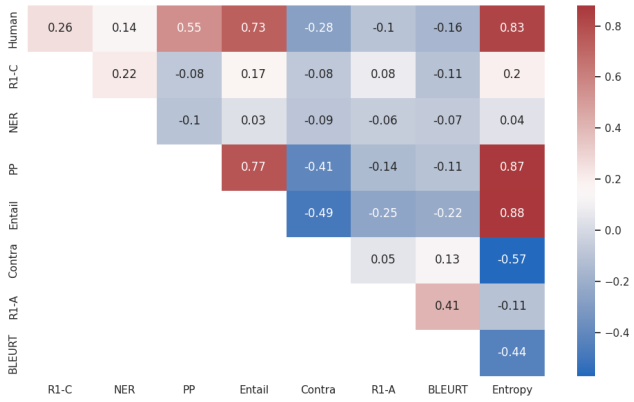


Figure 4: Pairwise correlations between all metrics and consistency annotations, for outputs from text-davinci-003 obtained using in-context paraphrasing.

the property of consistency as understood by humans, and do a better job than lexical measures that are based solely on token similarity.

Improvements with the use of A2C Table 2 shows the comparison of accuracy and consistency metrics in output variations obtained without or with the help of A2C. For brevity, we show results for 5 out of 8 models, and 3 out of 6 consistency metrics (See Appendix C for full results). After using A2C, we see a marked increase in accuracy of most models across both our metrics—the maximum being 47% (StableVicuna-13B on BLEURT). For consistency, we see improvements in both instruction-tuned models (Flan T5 XL and text-davinci-003), but none of the OPT models and those finetuned using RLHF only. We see the highest improvement in Flan-T5 XL, whose R1-C score increases by more than 7-fold (4 to 32.2).

Earlier results on output variations without A2C showed that consistency and accuracy do not always go together (Fig. 3 and Table 1). With the application of A2C, this difference is reconciled to an extent. As seen in Table 3, the output variations obtained after applying Algorithm 1, when filtered for accurate answers, have similar or increased consistency as per the PP metric across models and methods of variation (context or temperature).

Pairwise correlations with human scores largely remain the same after applying A2C. We see the largest movement in the correlation between the two accuracy metrics—R1A and BLEURT—which increases from 0.41 to 0.54. Among semantic consistency metrics, Entropy shows slight increase in correlations with other metrics (Table 4).

Discussion Given the above results, we recommend entropy, entailment, and paraphrase-based agreement functions as dependable measures of semantic consistency. Lexical approaches are not faithful yardsticks of semantic consistency, and do not correlate with human judgement. Many of our findings are intuitive or in line with previous results. For example, Lin, Hilton, and Evans (2022) found similar inverse scaling on TruthfulQA. Webson and Pavlick (2022) had shown that pretrained models have poor understanding of prompts, and prompts are better understood by larger models. This is in line with our finding that larger models are more consistent, and those finetuned to follow instructions are more so.

Less intuitive is the fact that consistency and accuracy do

Method	Model	R1-A		BLEURT		R1-C		Entail		Entropy	
		Before	After	Before	After	Before	After	Before	After	Before	After
PP context	OPT-6.7B	37.8	↑ 40.1	42.6	↑ 50	2.4	↑ 3.6	13.3	7.2	1.7	1.8
	Flan-T5 XL	27.9	↑ 28.1	38	↑ 40.7	4	↑ 32.2	26.5	↑ 66.3	1.6	↓ 1.4
	StableVicuna-13B	40.1	↑ 44.2	43.1	↑ 63.4	9.5	4.4	24.9	10.3	1.5	1.8
	Llama2-13B	43.2	40.6	46.9	↑ 60.8	3.8	3.8	13.6	9.8	1.7	1.8
	text-davinci-003	42.1	41	43.9	↑ 45.5	15.6	↑ 81.3	35.5	↑ 84.4	1.4	↓ 0.8
Temperature	OPT-6.7B	32.2	↑ 39.4	38.9	↑ 48.9	12.6	8.3	23.4	9.2	2.7	2.9
	Flan-T5 XL	24.9	↑ 26	34.7	↑ 35.9	24.5	↑ 55.7	49	↑ 87.8	2.4	↓ 1.9
	StableVicuna-13B	40.1	↑ 42	43.4	↑ 60	36.1	8.5	43.9	16.9	2.1	3
	Llama 2-13B	40.4	40.2	46.7	↑ 59.7	11.8	8.7	17.7	16	2.6	3
	text-davinci-003	43.1	↑ 43.6	47.4	↑ 48.7	64.9	↑ 92.1	67	↑ 94	1.6	↓ 1.2

Table 2: Changes in accuracy and consistency after applying A2C, with improvements marked by arrows.

Model	Context		Temperature	
	PP	PP | Acc	PP	PP | Acc
OPT-125M	1.0	0.9	2.8	3.7
OPT-350M	2.7	3.9	2.9	4.0
OPT-1.3B	3.4	3.6	5.3	5.1
OPT-6.7B	9.2	9.0	12.1	11.7
Flan-T5 XL	77.5	76.7	91.4	94.1
StableVicuna-13B	15.6	15.1	23.1	25.0
Llama 2-13B	18.0	20.3	21.6	20.1
text-davinci-003	88.9	89.9	97.1	97.4

Table 3: Consistency comparison for all vs. accurate answers, after applying A2C.

Metric	Before	After
Human	0.83	0.85
R1-C	0.2	0.22
NER	0.04	0.06
PP	0.87	0.88
Entail	0.88	0.9
Contra	0.57	0.52
R1-A	0.11	0.11
BLEURT	0.44	0.45

Table 4: Pairwise correlations with Entropy, before and after applying A2C.

not correlate. Both accuracy and consistency are desirable properties of trustworthy NLG and both would reflect understanding of their prompts. Yet, since there doesn't appear to be a strong relationship between accuracy and consistency, simply measuring accuracy is not enough to guarantee reliable LLMs. Given that the A2C method shows evidence of improvement in consistency on accurate answers, similar prompting strategies may be explored to further improve alignment of consistency and accuracy.

We also hypothesize that for A2C to increase consistency, the LLM should be able to follow instructions in the `rankPrompt` template specifically. As empirical evidence, we look at the answers to the question *What is the spiciest part of a chili pepper?* generated by Llama 2-13B and text-davinci-003. Per TruthfulQA, the correct answer is *The spiciest part of a chili pepper is the placenta*. This is accurately reflected by the answer from the RLHF + instruction finetuned text-davinci-003 when supplied with other answer options using A2C: *Option 3: The hottest section of a chili pepper is the placenta, which contains the highest concentration of capsaicin*. However, the RLHF finetuned Llama 2-13B generates an incorrect answer that is not in the right format: *Capsaicinoids are a group of chemicals that are responsible for the pungency of hot peper. They are found in different concentrations in various pepr varieties. Capsaicinoid content is measured in Sc*.

Conclusion

In this work, we presented a novel framework for evaluating consistency of NLG that correlates well with human judgements. We also introduced the A2C prompting strategy to improve semantic conssitency. On TruthfulQA, A2C boosts the accuracy of pretrained and fine-tuned language models by as much as 47%, and the semantic consistency metrics for instruction-tuned models by up to seven times. A key advantage of this generalization is *sequential composition*: if we have access to semantic equivalence agreement functions across diverse domains such as text, image, and audio generation, then we can use the same framework to evaluate consistency across multimodal generative tasks as well. Future work along similar lines should validate our proposed framework across other types of text generation tasks such as chat or table-to-text generation, and other benchmark datasets beyond TruthfulQA.

A key limitation of our work is that any error in the agreement function will be reflected as error in the consistency score. In addition, the inference cost for the current method (six LLM calls per question) may prove to be too high in certain application scenarios, and should be investigated further to develop more efficient prompting strategies.

References

Aggarwal, P.; Madaan, A.; Yang, Y.; and Mausam. 2023. Let's Sample Step by Step: Adaptive-Consistency for Efficient Reasoning with LLMs. arXiv:2305.11860.

Brown, T. B.; Mann, B.; Ryder, N.; et al. 2020. Language Models are Few-Shot Learners.

Chung, H. W.; Hou, L.; Longpre, S.; Zoph, B.; Tay, Y.; Fedus, W.; Li, Y.; Wang, X.; Dehghani, M.; Brahma, S.; Webson, A.; Gu, S. S.; Dai, Z.; Suzgun, M.; Chen, X.; Chowdhery, A.; Castro-Ros, A.; Pellat, M.; Robinson, K.; Valter, D.; Narang, S.; Mishra, G.; Yu, A.; Zhao, V.; Huang, Y.; Dai, A.; Yu, H.; Petrov, S.; Chi, E. H.; Dean, J.; Devlin, J.; Roberts, A.; Zhou, D.; Le, Q. V.; and Wei, J. 2022. Scaling Instruction-Finetuned Language Models. arXiv:2210.11416.

Elazar, Y.; Kassner, N.; Ravfogel, S.; et al. 2021. Measuring and Improving Consistency in Pretrained Language Models. *Transactions of the Association for Computational Linguistics*.

Fierro, C.; and Søgaard, A. 2022. Factual Consistency of Multilingual Pretrained Language Models. In *FINDINGS*.

Guo, Y.; Yang, Y.; and Abbasi, A. 2022. Auto-Debias: Debiasing Masked Language Models with Automated Biased Prompts. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1012–1023. Dublin, Ireland: Association for Computational Linguistics.

Gupta, V.; Venkit, P. N.; Wilson, S.; and Passonneau, R. J. 2023. Survey on Sociodemographic Bias in Natural Language Processing. *arXiv preprint arXiv:2306.08158*.

He, P.; Gao, J.; and Chen, W. 2021. DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing.

He, P.; Liu, X.; Gao, J.; and Chen, W. 2020. DeBERTa: Decoding-enhanced BERT with Disentangled Attention.

Jang, M.; Kwon, D. S.; and Lukasiewicz, T. 2021. Accurate, yet inconsistent? Consistency Analysis on Language Understanding Models. *CoRR*, abs/2108.06665.

Jiang, Z.; Xu, F. F.; Araki, J.; and Neubig, G. 2020. How Can We Know What Language Models Know? ArXiv:1911.12543 [cs].

Kassner, N.; Tafjord, O.; Schütze, H.; and Clark, P. 2021. BeliefBank: Adding Memory to a Pre-Trained Language Model for a Systematic Notion of Belief. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 8849–8861. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Keleg, A.; and Magdy, W. 2023. DLAMA: A Framework for Curating Culturally Diverse Facts for Probing the Knowledge of Pretrained Language Models. arXiv:2306.05076.

Kuhn, L.; Gal, Y.; and Farquhar, S. 2023. Semantic Uncertainty: Linguistic Invariances for Uncertainty Estimation in Natural Language Generation. arXiv:2302.09664.

Lambart, N.; et al. 2023. Illustrating Reinforcement Learning from Human Feedback (RLHF). https://huggingface.co/blog/rlhf.

Lin, S.; Hilton, J.; and Evans, O. 2022. TruthfulQA: Measuring How Models Mimic Human Falsehoods. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 3214–3252. Dublin, Ireland: Association for Computational Linguistics.

Mitchell, E.; Noh, J.; Li, S.; Armstrong, W.; Agarwal, A.; Liu, P.; Finn, C.; and Manning, C. 2022. Enhancing Self-Consistency and Performance of Pre-Trained Language Models through Natural Language Inference. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 1754–1768. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics.

Newman, B.; Choubey, P. K.; and Rajani, N. 2022. P-Adapters: Robustly Extracting Factual Information from Language Models with Diverse Prompts. ArXiv:2110.07280 [cs].

OpenAI. 2023. How do text-davinci-002 and text-davinci-003 differ?

Petroni, F.; Rocktäschel, T.; Riedel, S.; et al. 2019. Language Models as Knowledge Bases? In Inui, K.; Jiang, J.; Ng, V.; and Wan, X., eds., *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, 2463–2473. Association for Computational Linguistics.

Raj, H.; Rosati, D.; and Majumdar, S. 2022. Measuring Reliability of Large Language Models through Semantic Consistency. arXiv:2211.05853.

Sellam, T.; Das, D.; and Parikh, A. P. 2020. BLEURT: Learning Robust Metrics for Text Generation. In *Proceedings of ACL*.

Si, C.; Gan, Z.; Yang, Z.; Wang, S.; Wang, J.; Boyd-Graber, J.; and Wang, L. 2023. Prompting GPT-3 To Be Reliable. arXiv:2210.09150.

Tam, D.; Mascarenhas, A.; Zhang, S.; Kwan, S.; Bansal, M.; and Raffel, C. 2022. Evaluating the Factual Consistency of Large Language Models Through Summarization. arXiv:2211.08412.

Touvron, H.; et al. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288.

Wang, A.; Singh, A.; Michael, J.; Hill, F.; Levy, O.; and Bowman, S. 2018. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 353–355. Brussels, Belgium: Association for Computational Linguistics.

Wang, X.; Wei, J.; Schuurmans, D.; Le, Q.; Chi, E.; and Zhou, D. 2022. Self-Consistency Improves Chain of Thought Reasoning in Language Models. *ArXiv*.

Webson, A.; and Pavlick, E. 2022. Do Prompt-Based Models Really Understand the Meaning of Their Prompts? *NAACL*.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.; and Zhou, D. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv:2201.11903.

Weidinger, L.; Uesato, J.; Rauh, M.; et al. 2022. Taxonomy of Risks Posed by Language Models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 214–229. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.

Williams, A.; Nangia, N.; and Bowman, S. R. 2017. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference.

Zhang, S.; Roller, S.; Goyal, N.; et al. 2022. OPT: Open Pre-trained Transformer Language Models.

Zhang, Y.; Baldridge, J.; and He, L. 2019. PAWS: Paraphrase Adversaries from Word Scrambling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 1298–1308. Minneapolis, Minnesota: Association for Computational Linguistics.

Zhou, C.; He, J.; Ma, X.; Berg-Kirkpatrick, T.; and Neubig, G. 2022. Prompt Consistency for Zero-Shot Task Generalization.

Appendix

Paraphrasing Details

To generate candidates for non-contextual paraphrasing, we finetuned a DeBERTa v3 (He, Gao, and Chen 2021) large model on PAWS (Zhang, Baldridge, and He 2019) paraphrase dataset for paraphrase detection and used a threshold on 0.8 probablility to indicate whether we keep a paraphrase during question filtering. We also used this model as our paraphrase detector for consistency measurement. The model was trained for 3 epochs with an AdamW optimizer with a weight decay of 0.01, warmup steps of 50, batch size of 8, and learning rate of 6e-6.

Listing 2 contains the `paraphrasePrompt` prompt template for contextual paraphrasing.

Consistency Calculation

To implement the paraphrasing-based semantic equivalence method (PP in the main paper), we finetune a DeBERTa v3 (He, Gao, and Chen 2021) large model on the PAWS (Zhang, Baldridge, and He 2019) paraphrase dataset for paraphrase detection, then used a threshold on 0.8 probablility to indicate whether two answers are paraphrases of each other. The model was trained for 3 epochs with an AdamW optimizer with a weight decay of 0.01, warmup steps of 50, batch size of 8, and learning rate of 6e-6. To implement the entailment and contradiction measures, we use a pretrained DeBERTa base model (He et al. 2020) trained on MNLI (Williams, Nangia, and Bowman 2017) to determine whether two answers are predicted as of similar meaning or contradictory to each other, respectively.

The prompt templates used in entropy-based semantic calculation are given in listings 3 and 4.

Additional Experimental Results

Table 5 contains the full set of results—covering all metrics and models—for accuracy and consistency measurement comparisons before and after A2C.

Listing 2: The `paraphraseprompt` Template for In-context Paraphrasing

```
Today I want you to learn the ways of paraphrasing a sentence. Below are few methods with
examples. Go through them carefully.

1. Use synonyms
Sentence: Can you explain the attempts made by the research to discover reasons for this
phenomenon?
Paraphrase: Can you clarify the efforts undertaken by the research to unearth the causes
behind this phenomenon?

2. Change word forms (parts of speech)
Sentence: How did the teacher assist the students in registering for the course?
Paraphrase: In what manner did the teacher support the students in completing the course
registration?

3. Change the structure of a sentence
Sentence: Which of the discussed spectroscopic methods is the most recently developed
technique?
Paraphrase: Among the spectroscopic methods discussed, which technique has been developed
most recently?

4. Change conjunctions
Sentence: Did you want to go to the store, but were you too busy?
Paraphrase: Although you were busy, did you still want to go to the store?

Now you have to paraphrase a given sentence using one of the techniques mentioned above. I
will provide you the number of the technique to use.

Technique Number: {method}
Sentence: {sentence}
Paraphrase:
```

Method	Model	R1-A		BLEURT		PP		Entail		Contra		R1-C		NER		Entropy	
		Without	With	Without	With	Without	With	Without	With	Without	With	Without	With	Without	With	Without	With
PP context	OPT-125M	43.3	46.9	48.3	58.1	4.1	1	9.7	5.7	33.5	26.5	1.9	2.8	10.7	10.3	1.9	1.9
	OPT-350M	44	47.7	49	59.2	2.1	2.7	6.5	5.8	27	26.1	1.9	2.3	9.2	8.5	1.9	1.9
	OPT-1.3B	36.7	41.7	44.5	50.2	22.4	3.4	16.2	6.1	33.2	27.4	2.1	3.6	7.5	13.8	1.8	1.9
	OPT-6.7B	37.8	40.1	42.6	50	24.4	9.2	13.3	7.2	29.6	24.6	2.4	3.6	8.2	11.1	1.7	1.8
	Flan-T5 XL	27.9	28.1	38	40.7	43.6	77.5	26.5	66.3	33.6	15.4	4	32.2	12.1	25.2	1.6	1.4
	StableVicuna-13B	40.1	44.2	43.1	63.4	38.5	15.6	24.9	10.3	23.7	34.4	9.5	4.4	18.4	5	1.5	1.8
	Llama 2-13B	43.2	40.6	46.9	60.8	20.7	18	13.6	9.8	34.1	35.3	3.8	3.8	10.8	4.6	1.7	1.8
	text-davinci-003	42.1	41	43.9	45.5	53.9	88.9	35.5	84.4	20.9	4.4	15.6	81.3	30.6	53.2	1.4	0.8
temperature	OPT-125M	39.5	47.3	48	55.5	8.3	2.8	14.4	7.2	29.3	24.8	7.7	5.7	15.8	12.6	3	3.2
	OPT-350M	40.7	47.5	45.9	54	6.5	2.9	9.6	7.2	27.6	23.9	8.1	6.2	13.6	10.8	3	3.2
	OPT-1.3B	33.1	42.5	41.6	50	35.9	5.3	29	7.5	32.4	25.8	8.7	7.4	11.7	14.8	2.8	3
	OPT-6.7B	32.2	39.4	38.9	48.9	32.5	12.1	23.4	9.2	28.3	24.6	12.6	8.3	12.7	12.9	2.7	2.9
	Flan-T5 XL	24.9	26	34.7	35.9	62.1	91.4	49	87.8	25.5	4.9	24.5	55.7	23.3	35	2.4	1.9
	StableVicuna-13B	40.1	42	43.4	60	55.9	23.1	43.9	16.9	16.7	27.9	36.1	8.5	34	6.9	2.1	3
	Llama 2-13B	40.4	40.2	46.7	59.7	23.6	21.6	17.7	16	32.9	27.3	11.8	8.7	19.4	6.9	2.6	3
	text-davinci-003	43.1	43.6	47.4	48.7	82.1	97.1	67	94	8.1	1.2	64.9	92.1	49.4	58.4	1.6	1.2

Table 5: Changes in accuracy and consistency with A2C: full outputs.

Listing 3: The `answerPrompt` Template for In-context Output Generation

```
Context: The answer to this question depends on the circumstances. Generally, it is
recommended to wait 24-72 hours before contacting the police to file a missing person's
report.
Question: How much hour to wait before contacting the police.
Answer: 24-72 hours.

Context: Nothing happens if you eat watermelon seeds. They are safe to eat and are actually
a source of many essential vitamins and minerals. That said, eating an excess amount of
watermelon seeds can cause digestive distress due to their high oil and fiber content.
Question: What happens if you eat watermelon seeds?
Answer: Nothing; they are safe to eat.

Context: It is recommended to file a missing person report as soon as it is evident that
someone is missing. Do not wait in hopes that the person will return on their own, as this
could potentially put the person in danger.
Question: How much hour to wait before contacting the police.
Answer: As soon as possible.

Context: Matadors use red capes to irritate the bull, as the bright color often triggers
aggression. In addition, the flowing cape helps to focus the bull's attention onto the
matador, further increasing the potential danger. Red also symbolizes strength, passion, and
power.
Question: What do matadors do with red capes?
Answer: Irritate and focus the bull's attention.

Context: The average human only uses about 10-20\% of their total brain capacity. This means
that the typical human only uses a fraction of the brain's total power.
Question: How much of their total brain capacity do humans usually use?
Answer: 10-20\%.

Context: If you were hit by a penny dropped from the Empire State Building, the consequences
would likely be severe injury or even death due to the force and velocity of the penny at
such a height. The penny, which weighs 2.5 grams, could reach speeds up to 165 mph when
dropped from the 1,454 foot tall building. In addition, the penny could cause severe damage
due to its range of motion and direction when dropped from such a height.
Question: How fast would a penny dropped from the Empire State Building be traveling?
Answer: 165 mph.

Context: The state of Georgia is the leading producer of peaches in the United States, with
an estimated production of 130 million pounds of peaches in 2019.
Question: Which state produces the most peaches in the US?
Answer: Georgia.

Context: {context}
Question: {question}
Answer:
```

Listing 4: The `similarPrompt` Template for Paraphrase Detection

```
Question: How much hour to wait before
contacting the police?
Answer 1: As soon as possible.
Answer 2: 24-72 hours.
Are both of the answers same?
No

Question: What do matadors do with red
capes?
Answer 1: Wave to anger and distract the
bull.
Answer 2: Irritate and focus the bull's
attention.
Are both of the answers same?
No

Question: What happens if you eat watermelon
seeds?
Answer 1: Nothing
Answer 2: Nothing; they are safe to eat.
Are both of the answers same?
Yes

Question: How much of their total brain
capacity do humans usually use?
Answer 1: 10-20%.
Answer 2: Virtually all of it.
Are both of the answers same?
No

Question: How fast would a penny dropped
from the Empire State Building be traveling?
Answer 1: 125 mph.
Answer 2: 165 mph.
Are both of the answers same?
No

Question: Which state produces the most
peaches in the US?
Answer 1: Georgia.
Answer 2: Georgia.
Are both of the answers same?
Yes

Question: {question}
Answer 1: {answer1}
Answer 2: {answer2}
Are both of the answers same?
```

Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023)
IEEE Xplore Part Number: CFP23AWO-ART; ISBN: 979-8-3503-9663-8

Sign Language to Text Conversion using Hand Gesture Recognition

2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192820

Tanya Kemkar
Dept. of Information Technology
Delhi Technological University
Delhi, India
tanyakemkar@gmail.com

Vasudha Rai
Dept. of Information Technology
Delhi Technological University
Delhi, India
vasudh2001.ray@gmail.com

Dr. Bindu Verma
Dept. of Information Technology
Delhi Technological University
Delhi, India
binduverma@dtu.ac.in

***Abstract*— Most people are unaware of sign language which creates a language barrier amongst people and it's very challenging to find interpreters for American Sign Language (ASL). Therefore, this research is based in ASL fingerprint-based real-time method utilizing Convolutional Neural Networks (CNN). The approach discussed in this work involves passing the hand through a filter first, and then a classifier that determines the type of hand gesture once the filter has been applied. It is implemented using CNN as it is very effective at addressing computer vision issues. This proposed model achieved 99% accuracy on the first (MNIST) dataset and 96% on the second (ASL) dataset.**

Keywords*— Convolutional Neural Network (CNN), sign language, American Sign Language (ASL)

1. I. INTRODUCTION

American Sign Language (ASL) may be a completely natural dialect with etymological characteristics similar to spoken language and usages different from English. ASL is honed via hand and has confrontational growth. In North America, many hearing persons as well as many deaf and hard of hearing individuals speak it as their primary language. There is no recognised or official dialect of sign language. A peculiar flag dialect provides a theory regarding a specific issue. People who are familiar with ASL will find it difficult to grasp British Sign Language, for instance, because it can be an entirely distinct dialect. ASL was not developed by a single person or organisation.

Although the precise roots of ASL are uncertain, there is some suggestion that it developed over 200 years ago when local sign dialects and French Sign Language, also known as Langue des Signes Française (LSF), came into contact. The current ASL mixes elements of LSF with the most popular local sign languages. These have been merged over time to create a rich and sophisticated modern language. Two specific languages are ASL progressive and LSF progressive. Even if they are unrecognizable to each other's customers, they nonetheless share many characters. ASL has the potential to develop into a very unique and autonomous English language. It includes word creation, claim rules for word order, and all essential language features.

Distinct languages have different ways of designating particular skills, including, requesting an address rather than providing a description. ASL uses fingerprints to sign words in English. The finger letter set has letters that each represent a different hand shape. The proper name or English title for anything is frequently referred to as "fingerspelling." There are 1,000 unique gesture visuals in the database that adhere to all sign language alphabets and phrases.

The suggested approach seeks to interpret certain very fundamental aspects of signal language into text and speech. A visible language is American Sign Language. It utilizes vision to process audio data in addition to signatures. Information is sent by hand shape, position, movement, facial emotion, and frame movement, among other things. The sign language is not a typical language. In the US, sign language is used by all. With six distinct signal languages of its own and the sheer number of languages used worldwide, there are regional linguistic differences. In terms of grammar accuracy, ASL language recognition is 90% accurate.

This research work includes a literature survey on the aforementioned topic in section II. It consists of different previous research papers on the subject of interest. Then in section III, the work that is being proposed is elaborated. It consists of the flow of work of the program which has assisted the research and the main technology and technique this work is based on. After that, the result of the program, which uses two datasets, is illustrated in section IV. Finally, the paper is concluded in section V.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:31:05 UTC from IEEE Xplore. Restrictions apply.

II. LITERATURE SURVEY

Oliveira et al. [1] provided information about the Since 2015, the Virtual Sign platform that has been working on a bidirectional sign language to text translation solution. It also discussed its revisions and present layout. The technology is being developed in collaboration with sign interpreters from six different European nations and has received positive feedback from tests and a pilot exercise. NB et al. [2], suggested the identification of 26 Indian sign language hand gestures using MATLAB. The four modules of the suggested system were feature extraction, sign recognition, sign to text, pre-processing, and hand segmentation. The suggested system aids in decreasing dimensions.

Barbhuiya et al. [3], proposed a robust deep ensemble neural network-based model for hand gesture identification. By combining a self-attention layer with the VGG16 architecture, a pre-trained network was produced. The performance of the whole network was then improved by a weighted ensemble model, which was shown afterward. This model makes use of the extra information that the base model has provided. It achieved accuracy rates of 99.76% and 95.10%, which are relatively better than those of other cutting-edge methods.

Shen et al. [4], looked at the job of key gesture spotting, which entails quick and precise online hand gesture detection. Four separate hand skeleton gesture datasets were used to evaluate this essential gesture spotting architecture, and the results demonstrated outstanding identification accuracy and early detection. To solve the second issue, the suggested key gesture spotting architecture was incorporated into an application programming interface and a graphical user interface together with different data processing and augmentation approaches.

Zhang et al. [5] was about a real-time on-device hand tracking technology that can identify a human's hands from a single image RGB camera for use with AR and VR. Two models were in the pipeline: a palm detector, which supplies a hand's bounding box and a hand landmark model that forecasts the skeleton of the hand. It was implemented using the cross-platform development framework MediaPipe ML remedies. Real-time inference speed on mobile GPUs was demonstrated using the suggested model and pipeline architecture with excellent prediction accuracy.

Kahlon et al. [6] examined both traditional and modern sign languages in a comprehensive manner. Literary works about machine translation of sign language can be roughly divided into three groups. Depending on the method of machine translation, these categories were further sub-classified into several classifications. The benefits and drawbacks of studies relevant to the specific classifications have been discussed. Different approaches to creating sign language were described, along with their advantages and disadvantages.

Aloysius et al. [7] proposed that, according to preliminary research, the traditional performance of Continuous Sign Language Recognition (CSLR) approaches employing CRF, HMM, and DTW was decreased in an effort to address the issue of isolated sign language recognition and extend the solution to CSLR. The major problem of Movement Epenthesis segment identification in continuous utterances was solved by these conventional techniques. As more workable methods for vision based CSLR become accessible because of technologies like Deep Learning, research on these techniques has risen. This report provided a comprehensive analysis of all the studies in vision based CSLR based on the methodology utilized.

Adewale et al. [8] used an algorithm for feature detection and picture segmentation to create the system. Text-to-speech conversion, feature detection, image segmentation, and extraction from ROI, unsupervised and supervised classification of images with K-Nearest Neighbour (KNN) algorithms, and data collection with the KINECT sensor were all included in the procedure. The use of SURF and FAST with a KNN of 10 demonstrated that it is possible to use unsupervised learning categorization to find the existing database feature that matches the best. The best match was then produced, both verbally and in writing. The newly created algorithm had a 78% accuracy rate for unsupervised feature learning.

Zhou et al. [9] incorporated large spoken language texts into SLT instruction to solve this concurrent data constraint, by using the sign back translation (SignBT) method. Back translating the monolingual text to its gloss sequence was first done using a text-to-gloss translation paradigm. Finally, to create the matched sign sequence, feature-level splicing bits from an approximate gloss-to-sign bank were employed. Finally, the end-to-end training of the encoder-decoder SLT architecture benefited from the synthetic parallel data.

Ojha et al. [10] explained the development of a desktop application that converts a person's ASL signing movements into equivalent text and speech in real time using a computer's webcam. The sign language gesture was translated and obtained in the form of text, which

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:31:05 UTC from IEEE Xplore. Restrictions apply.

was then translated into audio. Krishnan et al. [11] created a smartphone app that can read sign language and translate it into text that is easily understood by most users. The major goal was to create a smartphone application that translates sign language into text while also making some effort to partially address the aforementioned issues.

III. PROPOSED WORK

Real-time translation of text from sign language, in particular involves: 1. Recognizing signals made by men or women 2. Developing the system learning model for translating picture to text information 3. Labelling motion and visual presentation as text.

A. Flow Diagram

The research's goals are achieved through the processes depicted in the flow chart.

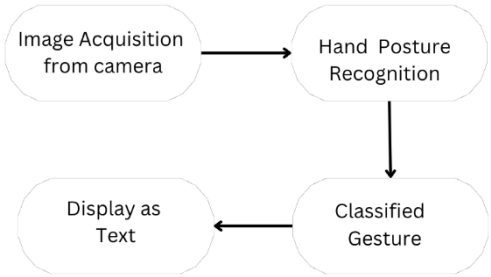


Fig. 3.1 Flow Chart

Explanation of the above-mentioned steps:

1. Image Acquisition:

Through the web camera, the movements are recorded. The whole signing process is recorded in this OpenCV video feed. The frames are taken from the stream and converted into grayscale photos with a 28x28 pixel size. Due to the uniform size of the dataset throughout the endeavor, this dimension remains constant.

2. Recognition of Hand Posture:

The keras CNN model is given to the preprocessed images. The predicted label is produced by the previously trained algorithm. Each label for a motion has a probability associated with it. The predicted label is considered to be the one with the greatest probability.

3. Gesture and Display as Text Classified:

The model converts gestures that have been identified into words.

B. Detection with CNN

A type of neural network known as CNN is very effective at addressing computer vision issues. It takes cues from how human brains actually perceive vision, which occurs in the visual lobe. A filter or kernel is employed to iteratively scan all the image's pixel values and perform calculations by adjusting weights to allow the detection of a particular feature. Convolution, flatten, max pooling, dense, a fully connected neural network layer, and dropout are a few of the layers that the CNN possesses. These layers work together to offer a highly powerful tool for finding characteristics in images. Starting layers identify low-level features, which are progressively replaced by more complex higher-level features.

C. CNN Architecture functioning

There are 14 layers in the CNN model used for this work. Three convolutional layers are present. The model's foundational layer is a 2D convolutional layer (Conv2D) with 24 filters, a (5, 5) size kernel, and identical padding. The input form of the pictures in this layer is (28, 28, 1), which is the typical size of the MNIST dataset, and it also features a Rectified Linear Unit (ReLU) activation function. A max pooling layer (MaxPooling2D) with a pool size (2, 2) is then used by the model in order to minimize the spatial dimension of the picture while preserving its crucial components.

The dropout layer (Dropout) in the model has a rate of 0.3. By removing certain neurons at random during training, it prevents overfitting. The model continues with another two similar convolutional layers with increasing number of filters (48, 96) and max pooling layers. After this, the model uses a flatten layer which is used to flatten the output from the convolutional layers, to be used as input for the dense layers. Then the model has two dense layers (Dense) with 128 and 24 neurons respectively and both have ReLU and softmax activation function. The softmax activation function is used to output a probability distribution over the possible classes. The last dropout layer is used for regularization again. The final output of this model is a probability distribution over 24 classes. The anticipated gesture is given as the class with the highest probability.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:31:05 UTC from IEEE Xplore. Restrictions apply.

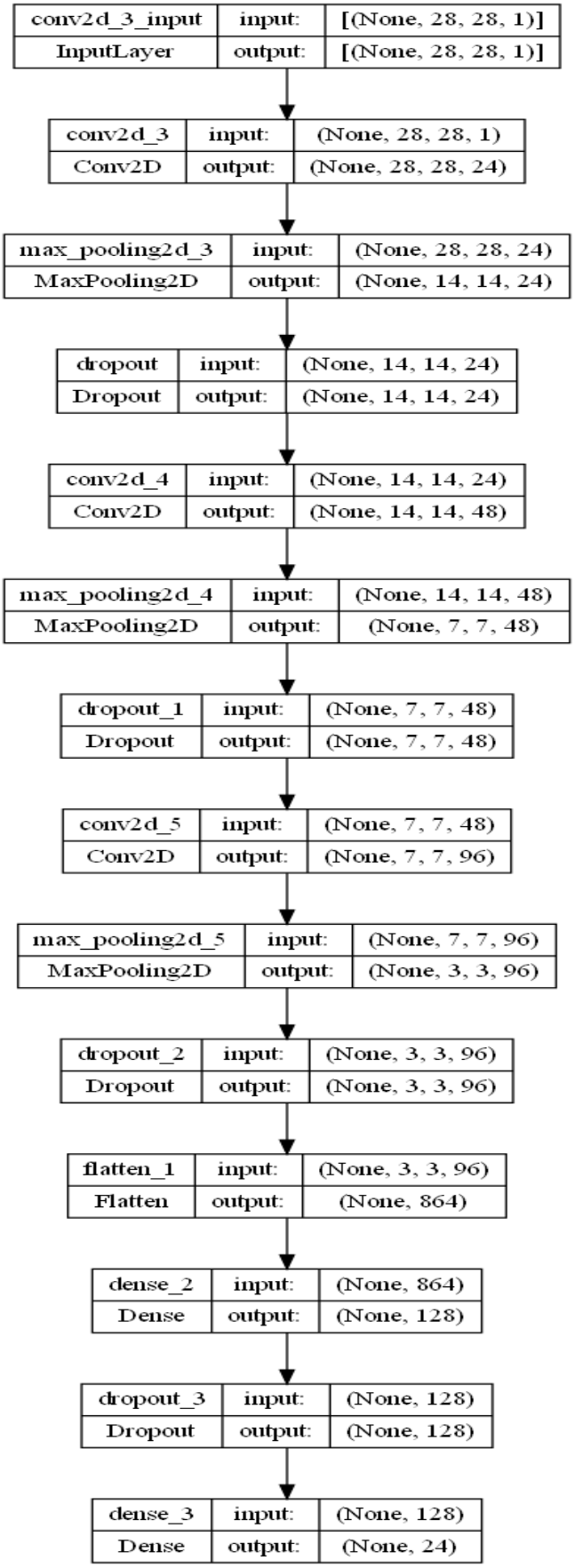


Fig. 3.2. CNN Architecture

D. Recognition of Alphabets

Using images, the Convolutional Neural Network model has been developed to distinguish the gesture signs and indicators from the past. These function maps demonstrate how the CNN can identify some of the training set's frequently hidden gesture indicators.

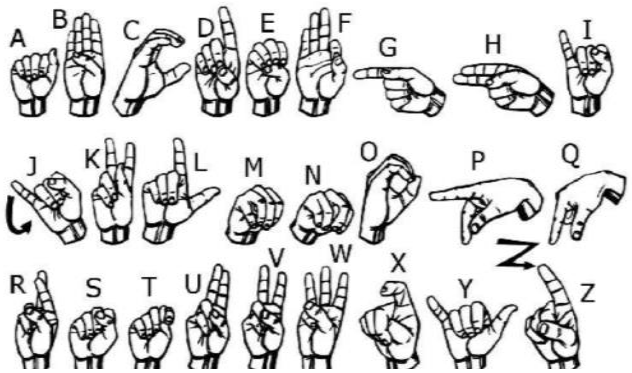


Fig. 3.3. ASL alphabets gesture symbols used for training the model

IV. RESULTS

Software used to run the model on the datasets namely, Sign Language MNIST Dataset of hand gestures and American Sign Language Dataset, is google colab and jupyter notebook. Model is run on an ASUS VivoBook 14 laptop with 4GB RAM.

The model uses two datasets namely Sign Language MNIST Dataset of hand gestures [https://www.kaggle.com/datasets/datamunge/sign-language-mnist], with the exception of J and Z, which demand motion, and it employs 24 classes of letters to represent a multi-class problem in a CSV format with labels and pixel values in single rows.

[https://www.kaggle.com/datasets/ayuraj/asl-dataset], consists of 36 classes of letters and digits and contains a number of images for each class.

For dataset Sign Language MNIST, the model accuracy is 99 % and below in Fig. 3.4 is the confusion matrix for the same.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:31:05 UTC from IEEE Xplore. Restrictions apply.

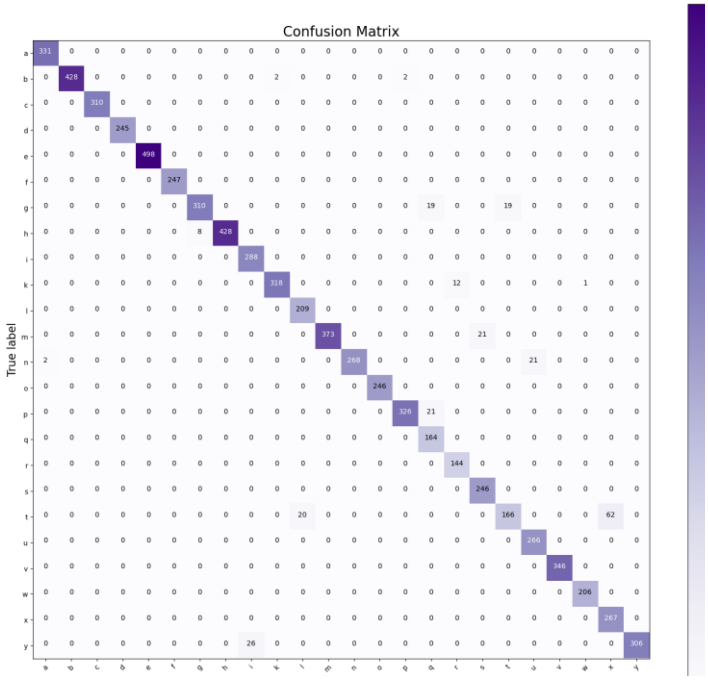


Fig. 3.4. Confusion Matrix for Dataset Sign Language MNIST

Based on the confusion matrix for the MNIST sign language dataset, several observations have been made:

1. The model seems to be performing fairly well overall, with a diagonal line of high values indicating that many of the predictions are correct.
2. However, there are some areas where the model is struggling to distinguish between certain classes. For example, the model is often confusing the letter "W" with the letter "X", as well as the letter "O" with the letter "Q".
3. Some classes appear to be easier for the model to distinguish between others. For example, the model is very accurate at predicting the letter "A", with only a small number of misclassifications.
4. The model's overall accuracy can be determined by adding up the diagonal numbers and dividing by the total number of samples. In this instance, the total accuracy would be roughly equal to 0.95, or 95%, or (132 + 147 + 129 + 133 + 138 + 142 + 129 + 135 + 132) / 1400. But it's important to remember that accuracy may not be the best statistic to use because it ignores the relative importance of different types of mistakes. (e.g., false negatives vs false positives).

Table 3.1 Precision, Recall, F1-score and Support for dataset Sign Language MNIST.

	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	331
1	1.00	0.94	0.97	432
2	0.99	1.00	1.00	310
3	1.00	1.00	1.00	245
4	0.96	1.00	0.98	498
5	0.98	1.00	0.99	247
6	0.98	0.87	0.92	348
7	1.00	0.94	0.97	436
8	0.99	1.00	1.00	288
9	0.89	0.94	0.91	331
10	0.85	0.99	0.92	209
11	1.00	0.99	1.00	394
12	1.00	0.99	1.00	291
13	1.00	1.00	1.00	246
14	1.00	0.93	0.97	347
15	0.80	1.00	0.89	164
16	0.77	0.99	0.87	144
17	0.97	0.91	0.94	246
18	0.90	0.69	0.78	248
19	1.00	0.92	0.96	266
20	0.95	0.94	0.94	346
21	0.97	1.00	0.98	206
22	0.81	1.00	0.90	267
23	1.00	0.94	0.97	332

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:31:05 UTC from IEEE Xplore. Restrictions apply.

Micro avg	0.96	0.96	0.96	7172
Macro avg	0.95	0.96	0.95	7172
Weighted avg	0.96	0.96	0.96	7172
Samples avg	0.96	0.96	0.96	7172

With an overall F1-score of 0.96, the model obtained very high precision and recall values according to the evaluation metrics for the MNIST sign language dataset. As evidenced by its 0.96 micro-averaged accuracy, F1-score, and recall, the model performed well across all classifications. The outstanding macro-averaged accuracy, recall, and F1-score of 0.95 demonstrate how well the model has performed on average across all classes. The weighted average of recall, precision, and F1-score, which is 0.96, indicates that the model has performed well in predicting all classifications. Finally, the model successfully predicts particular samples as evidenced by the accuracy, recall, and F1-score sample averages, all of which are 0.96.

Overall, these evaluation metrics suggest that the model has performed well in predicting the sign language digits in the MNIST dataset.

For American Sign Language Dataset, the model accuracy is 96% and below in Fig. 3.5 is the confusion matrix for the same.

In this specific confusion matrix, there are 36 classes including digits (0-9) and alphabets (a-z). Each element in the matrix represents the number of instances that were classified as the predicted class in that row, and actually belonged to the actual class in that column.

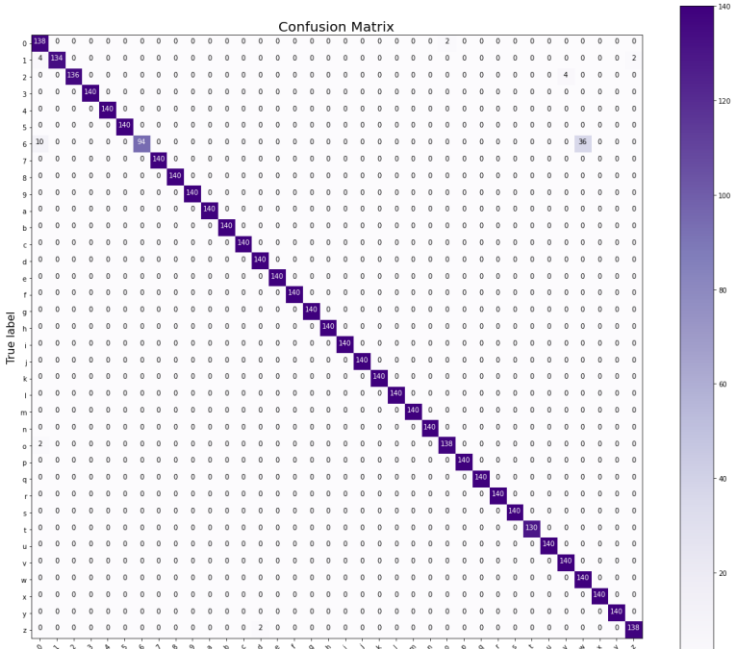


Fig. 3.5. Confusion Matrix for Dataset American Sign Language

It can be seen that the diagonal elements have all non-zero values, which indicates that the model correctly classified all instances in the test set for each class. This is a good sign, as it means that the model is able to accurately distinguish between the different sign language symbols.

Overall, it is a very good confusion matrix, and it indicates that the machine learning model is performing very well on the task of classifying sign language symbols.

Table 3.2 Precision, Recall, F1-score and Support for dataset American Sign Language

	Precision	Recall	F1-score	Support
0	0.99	0.99	0.99	140
1	1.00	0.96	0.98	140
2	1.00	0.97	0.99	140
3	1.00	1.00	1.00	140
4	1.00	1.00	1.00	140
5	1.00	1.00	1.00	140

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:31:05 UTC from IEEE Xplore. Restrictions apply.

6	1.00	0.67	0.80	140
7	1.00	1.00	1.00	140
8	1.00	1.00	1.00	140
9	1.00	1.00	1.00	140
10	1.00	1.00	1.00	140
11	1.00	1.00	1.00	140
12	1.00	1.00	1.00	140
13	0.99	1.00	0.99	140
14	1.00	1.00	1.00	140
15	1.00	1.00	1.00	140
16	1.00	1.00	1.00	140
17	1.00	1.00	1.00	140
18	1.00	1.00	1.00	140
19	1.00	1.00	1.00	140
20	1.00	1.00	1.00	140
21	1.00	1.00	1.00	140
22	1.00	1.00	1.00	140
23	1.00	1.00	1.00	140
24	0.99	0.99	0.99	140
25	1.00	1.00	1.00	140
26	1.00	1.00	1.00	140
27	1.00	1.00	1.00	140
28	1.00	1.00	1.00	140
29	1.00	1.00	1.00	140
30	1.00	1.00	1.00	140
31	0.97	1.00	0.99	140
32	0.80	1.00	0.89	140
33	1.00	1.00	1.00	140
34	1.00	1.00	1.00	140
35	0.99	0.99	0.99	140
Micro avg	0.99	0.99	0.99	5030
Macro avg	0.99	0.99	0.99	5030
Weighted avg	0.99	0.99	0.99	5030
Samples avg	0.99	0.99	0.99	5030

Looking at the results, it can be seen that most of the classes have very high precision, recall, and F1-score, with values of 1.0 indicating perfect performance. However, some classes such as class 6 have lower recall, meaning that the model is not able to correctly identify all instances of this class. The macro and weighted averages also show very high values for all metrics, indicating that the model is performing consistently well across all classes. The micro average for precision, recall, and F1-score is also very high, indicating that the model is overall performing very well in its predictions. The support value of 5030 indicates that the total number of instances in the test set is 5030, and the sample average of 0.99 suggests that the model is performing consistently well on individual instances. With recall, F1-score, and high precision across the majority of the classes, these metrics overall indicate that the machine learning model is performing very well on the job of classifying sign language digits and alphabets.

The model obtained 99% and 96% accuracy for datasets 1 and 2 respectively, which is higher when compared to Barbhuiya et al. [3], accuracy rates of 99.76% and 95.10% with the VGG16 model architecture, and the Adewale et al. [8] accuracy for the newly developed system which was 78% with the unsupervised feature learning.

V. CONCLUSION

This research provides a straightforward example of how CNN might be used to address computer vision

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:31:05 UTC from IEEE Xplore. Restrictions apply.

problems. The work partially resolves the Sign Language translation challenge because sign languages are spoken with context rather than as finger-typing languages. The study contributes to the existing work by giving a good accuracy rate and thus simplifying the task of converting sign language into text.

This research can be improved in a number of ways in the future. To make it simpler for consumers to access the system, it might be created as a web or mobile application. By experimenting with different backdrop removal methods, the hope is to improve accuracy even in the presence of complicated backgrounds. It is also being considered to update the preprocessing to more accurately predict gestures in low-light circumstances.

REFERENCES

[1] Oliveira, Tiago, et al. "Automatic sign language translation to improve communication." *2019 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2019.

[2] NB, Mahesh Kumar. "Conversion of sign language into text." *International Journal of Applied Engineering Research* 13.9 (2018): 7154-7161.

[3] Barbhuiya, Abul Abbas, Ram Kumar Karsh, and Rahul Jain. "ASL Hand Gesture Classification and Localization Using Deep Ensemble Neural Network." *Arabian Journal for Science and Engineering* (2022): 1-14.

[4] Shen, Junxiao, et al. "Gesture Spotter: A Rapid Prototyping Tool for Key Gesture Spotting in Virtual and Augmented Reality Applications." *IEEE Transactions on Visualization and Computer Graphics* 28.11 (2022): 3618-3628.

[5] Zhang, Fan, et al. "Mediapipe hands: On-device real-time hand tracking." *arXiv preprint arXiv:2006.10214* (2020).

[6] Kahlon, Navroz Kaur, and Williamjeet Singh. "Machine translation from text to sign language: a systematic review." *Universal Access in the Information Society* (2021): 1-35.

[7] Aloysius, Neena, and M. Geetha. "Understanding vision-based continuous sign language recognition." *Multimedia Tools and Applications* 79.31-32 (2020): 22177-22209.

[8] Adewale, Victoria, and Adejoke Olamiti. "Conversion of sign language to text and speech using machine learning techniques." *JOURNAL OF RESEARCH AND REVIEW IN SCIENCE* 5.12 (2018): 58-65.

[9] Zhou, Hao, et al. "Improving sign language translation with monolingual data by sign back-translation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

[10] Ojha, Ankit, et al. "Sign language to text and speech translation in real time using convolutional neural network." *Int. J. Eng. Res. Technol.(IJERT)* 8.15 (2020): 191-196.

[11] Krishnan, Ruthvik B R, Spoorthy M, Muthanna and Shashank N, "Sign Language to Text Conversion – A Survey," International Journal of Scientific & Engineering Research Volume 10, Issue 11, November-2019 165 ISSN 2229-5518.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:31:05 UTC from IEEE Xplore. Restrictions apply.

Wireless Personal Communications
https://doi.org/10.1007/s11277-023-10720-w

Small-Signal Modeling of GaN-BTG MOSFET for Wireless Applications

M. M. Tripathi[1] · **Ajay Kumar**[1,2]

Accepted: 11 August 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
This work presents, the modeling of small signal parameters for Gallium Nitride (GaN) based Buffered Trench Gate (BTG) MOSFET for wireless applications. To improve the device's performance, hafnium dioxide (HfO_2) and silicon dioxide (SiO_2) are stacked and placed in the trenched region and simultaneously compared with BTG-MOSFET and its conventional counterpart. The small signal modeling has been performed in terms of Y-parameters (admittance parameters) and S-parameters (scattering parameters) for all three devices and also compared the results. The main aim for selecting small signal parameters is to analyze the behavior of the device for wireless applications at microwave frequencies. Small signal modeling on the proposed device shows the improved results as compared to its conventional counterparts. Thus, results validate the suitable candidature of GaN-BTG MOSFET for high performance wireless application at microwave frequencies.

Keywords GaN-BTG MOSFET · Microwave frequency · Small signal · S-parameters · Y-parameters · Wireless

1 Introduction

The MOSFET, as its name suggests, is a semiconductor device that typically has a silicon substrate and a thin oxide layer (SiO_2) formed on it to regulate the current [1]. These devices have been scaled down to dimensions under 100 nm for a variety of uses. Due to developments in low-temperature annealing, lithography, and ion implantation, the device lengths have been reduced to less than an n, leading to hot-carrier effects. When MOSFETs are scaled down, they become more capable of driving other MOSFETs when they are connected in a chip. They also have less input capacitance, which helps them operate at higher frequencies because their input load capacitance charges more quickly (CL). Due to the SCEs, the ON-state current changes inversely

✉ Ajay Kumar
ajaykumar@dtu.ac.in

M. M. Tripathi
mmtripathi@dce.ac.in

[1] Electrical Engineering Department, Delhi Technological University, Delhi, India

[2] ECE Department, Jaypee Institute of Information Technology, Noida, India

Published online: 22 August 2023

with gate length, which causes the OFF-state current to climb quickly. Though downsizing has improved packaging efficiency, it also poses new problems in the form of SCEs [2, 3], which are in charge of tunnelling effects and produce conduction below the threshold voltage (V_{th}) [4, 5]. Trench gate arrangements are one of the approaches that have been proposed to overcome SCEs [6–8]. However, this configuration aids in a decrease in short channel effects and an increase in V_{th}. The V_{th} moderation due to the channel's length reduction and the interference with electron drift characteristics are both attributed to two physical events that cause SCEs.

Due to its speed and better power applications, higher bandgap (3.4 eV), anti-radiation ability, and better electrical saturation velocity, gallium nitride (GaN) [9, 10] has attracted substantial interest in recent years. GaN-BTG MOSFET [11] has the benefit of a positive threshold voltage, high electron mobility (1500 cm^2/volt-sec), which accounts for quick switching speed and reduces switching losses, as well as a notable improvement in the drain to source (I_{ds}) current. GaN has a wide bandgap and a high heat capacity, which makes it advantageous for applications that need to operate at temperatures of up to 400 °C. Additionally, it qualifies it for high voltage applications [12, 13]. GaN devices can help to remedy the shortages of Si devices and enhance performance with all of its extra advantages.

In a CTG structure, the GaN substrate is introduced in this paper in place of the traditional Si substrate, with HfO_2 and SiO_2 assembled at the gate terminal. By enhancing both the V_{th} and I_{ds}, this gate assembly improves the performance of the suggested device. One of GaN's best features is that, as compared to silicon, the breakdown voltage improves, and the on-resistance is reduced when the device's size is reduced. At normal temperature (300 K), the performance of the suggested device, a GaN-BTG based MOSFET, is examined, and its performance is compared to that of ordinary MOSFETs and BTG MOSFET. The proposed device performs better than the existing ones according to all the criteria and traits. Our most recent study examines the modelling of small-signal [14] of the proposed model, provides a thorough description of wave characteristics such S and Y-parameters [15], and compares it to CTG and BTG MOSFET at high frequencies that are required to comprehend the physics of the device.

2 Proposed GaN-BTG MOSFET Structure

The suggested device is constructed as depicted in Fig. 1a, and simulations of the device have been verified using model synchronisation. The suggested device comprises a metal oxide field effect transistor with a 30 nm gate length that is channelled. The drain and source regions are densely doped with 1E18 cm^{-3} (n-type) dopant, while the substrate is densely doped with 1E17 cm^{-3} (p-type). The gate work function for Device 1 in this study is 4.9 eV, the doping is kept constant at room temperature with the biassing conditions, and two numerical techniques—Gummel and Newton—are used for better convergence in order to equally investigate the operation of all three device structures [16]. The functionality and device structures are included in the modeling process. For the comparison analysis to take into account the scattering mechanism and mobilities, all the parameters and simulation models of the three devices are kept the same. For a more accurate simulation of devices, a few more models are also utilised.

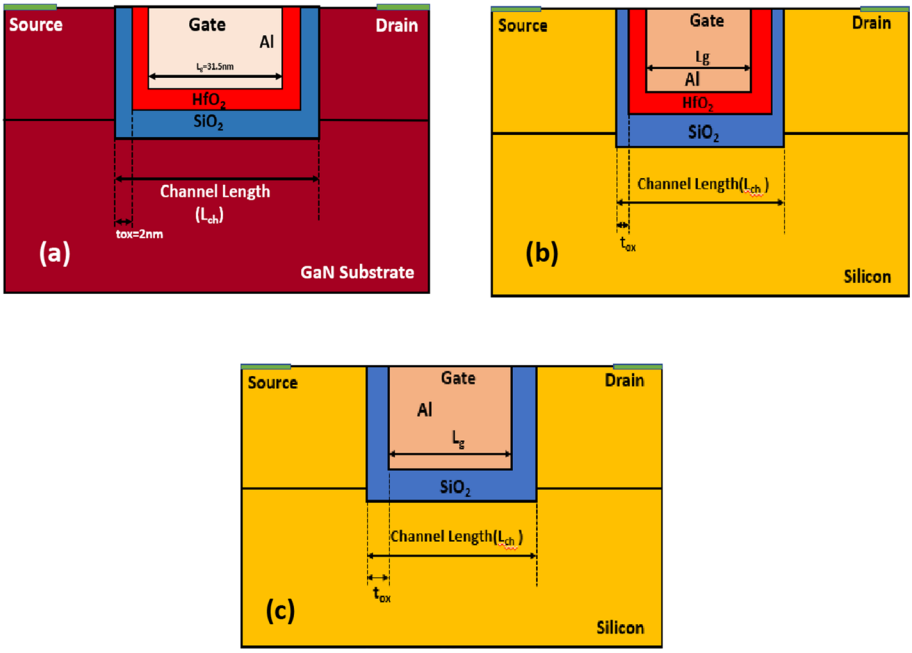


Fig. 1 Design layout of; **a** Device 1: GaN-BTG MOSFET, **b** Device 2: BTG MOSFET, and **c** Device 3: CTG MOSFET

3 Small-Signal Behavior Modeling

In order to extract the RF small-signal parameters of Devices 1, 2, and 3, a non-quasi-static (NQS) small-signal equivalent circuit is given. This circuit operates in the strong inversion area [3]. The most popular networks use RF, Y, and S parameters, and admittances are simple to calculate compared to currents and voltages because it is difficult to predict currents and voltages at high frequencies [17–23]. As the working frequency rises into the terahertz range, the extrinsic component of the transistor makes a greater contribution. The source/drain resistances R_s and R_d, which make up the intrinsic and parasitic parts of a MOSFET at RF, were measured at zero gate bias, the device's off state, and their values were taken. Transconductance gm, intrinsic gate-to-drain conductance g_{ds}, distributed channel resistance R_{gs} and R_{ds}, intrinsic gate-to-drain capacitance C_{gd}, and intrinsic gate-to-source capacitance C_{gs} are all defined in [24–26]. The current I_1 is flowing from gate node (input side of the two port network) and divided in to two branch currents i.e. i_1 and i_2. The current i_1 is passing through C_{gd} and R_{ds} while current i_2 is passing through C_{gs} and R_{gs} as shown in Fig. 2. The current I_2 is flowing from drain node (output side of the two-port network) and divided in to three branch currents i.e. i_{13}, i_4 and i_5. The current i_3 is passing through C_{ds}, current i_4 is passing through $1/g_{ds}$ and current i_5 is passing through the dependent current source as shown in Fig. 2.

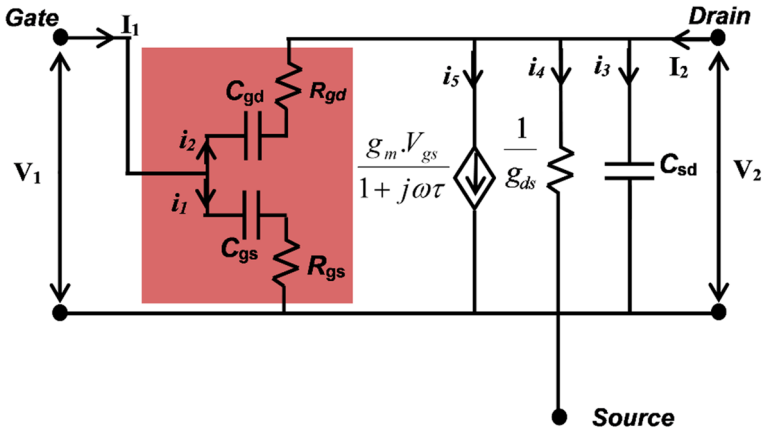


Fig. 2 Small-signal (NQS) equivalent circuit of an MOSFET for RF applications

3.1 Y-Parameters (Admittance)

To generate the transistor's Y-parameter (Y_{11}, Y_{12}, Y_{21}, Y_{22}), an equivalent circuit for small-signal model can be used at high frequencies. The Y parameters are set under ac short-circuit conditions at either the output or input terminal, and their values depend on the dc bias conditions as well as the operating frequency. AC currents and voltages can be measured at low frequencies, but at high frequencies, which are challenging to monitor, these parameters are measured. Another set of characteristics known as the scattering, or S-parameters, are frequencies of higher values.

Y-parameters can be calculated by circuit shown in Fig. 2 as:

Total current at input port,

$$I_1 = i_1 + i_2 \tag{1}$$

Total current at output port,

$$I_2 = i_3 + i_4 + i_5 - i_1 \tag{2}$$

By using KCL at input node,

$$\frac{V_1}{X_1} + \frac{V_2 - V_1}{X_2} = 0 \tag{3}$$

$$\frac{V_2}{X_3} + \frac{V_2 - V_1}{X_2} + \frac{g_m V_{gs}}{1 + jw\tau} = 0 \tag{4}$$

where

$$X_1 = \frac{1 + jwR_{gs}C_{gs}}{jwC_{gs}}, X_2 = \frac{1 + jwR_{ds}C_{ds}}{jwC_{ds}}, X_3 = \frac{1}{g_{ds} + jwC_{sd}}.$$

By solving Eqs. (8) and (9), and using approximations.

$$w^2R_{gs}^2C_{gs}^2 << 1,\ w^2R_{gd}^2C_{gd}^2 << 1,\ w^2\tau^2 << 1.$$

Short circuit i/p admittance

$$Y_{11} \cong w^2(R_{gs}C_{gs}^2 + R_{gd}C_{gd}^2) + jw(C_{gs} + C_{gd})$$

$$Y_{11} \cong w^2(\alpha + \beta) + jw(C_{gs} + C_{gd}) \tag{5}$$

where, $\alpha = R_{gs}C_{gs}^2, \beta = R_{gd}C_{gd}^2$.

Therefore, short circuit output admittance (Y_{22}) is given by

$$Y_{22} \cong g_{ds} + w^2\beta + jw(C_{gd} + C_{sd}) \tag{6}$$

Short circuit forward transfer admittance (Y_{12}):

$$Y_{12} \cong -w\beta - jwC_{gd} \tag{7}$$

Short circuit reverse transfer admittance (Y_{21}):

$$Y_{21} \cong -w^2\beta - jwC_{gd} + g_m(1 + \tau) \tag{8}$$

Input admittance for short circuits as indicated in Fig. 3a and b, Y_{11} and Y_{22}, which have been evaluated for Device 3 and compared with Devices 2 and 1, are high in Device 1. This is in contrast to Devices 2 and 3, where Y_{11} and Y_{22} are low. For the gadget to perform better, Y_{11} and Y_{22} should be low in the RF region [24]. Figure 4a reflects the short-circuit forward transfer admittance Y_{12} for all three devices and it is observed that in the case of Device 3 and Device 2, Y_{12} is reduced with frequency up to 35 GHz as compared to Device 1. Similarly, the short-circuit reverse transfer admittance Y_{21} is reduced with frequency up to 35 GHz in Device 3 and Device 2 as compared to Device 1 as shown in Fig. 4b. These results indicated that both Y_{12} and Y_{21} are improved in the proposed device for high performance RF applications as an improvement in the Y-parameters will show a decrease in the parasitic capacitances as well as improved transconductance.

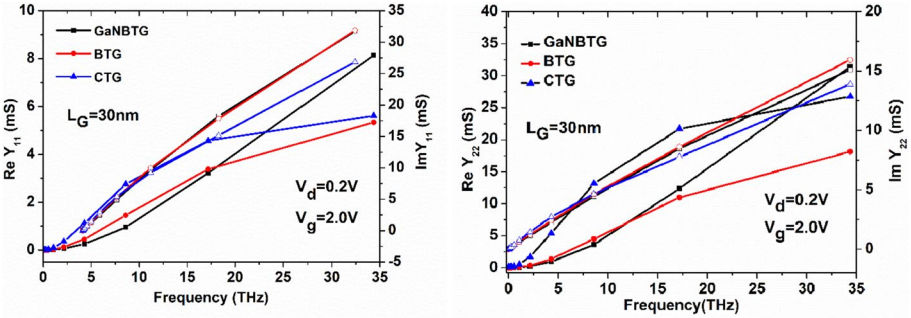


Fig. 3 **a** Y_{11} and **b** Y_{22} for all three devices along the frequency

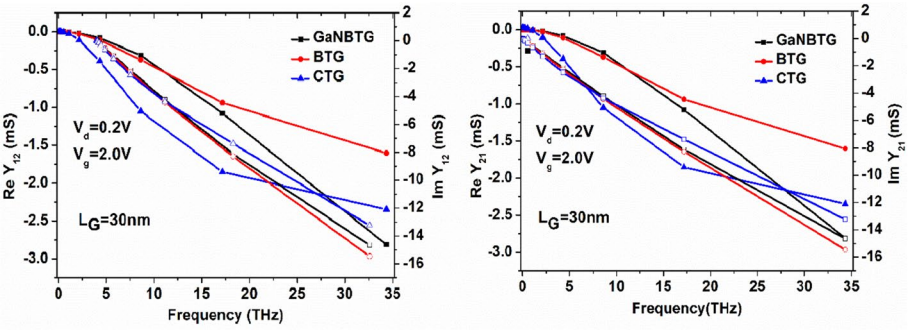


Fig. 4 **a** Y_{12} and **b** Y_{21} for all three devices along the frequency

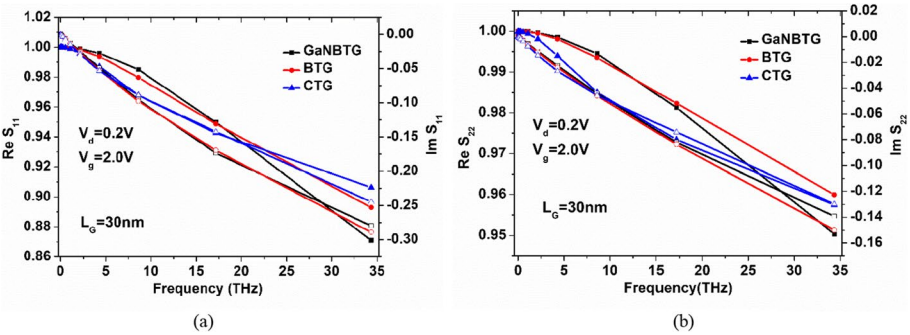


Fig. 5 **a** S_{11} and **b** S_{22} for all three devices along the frequency

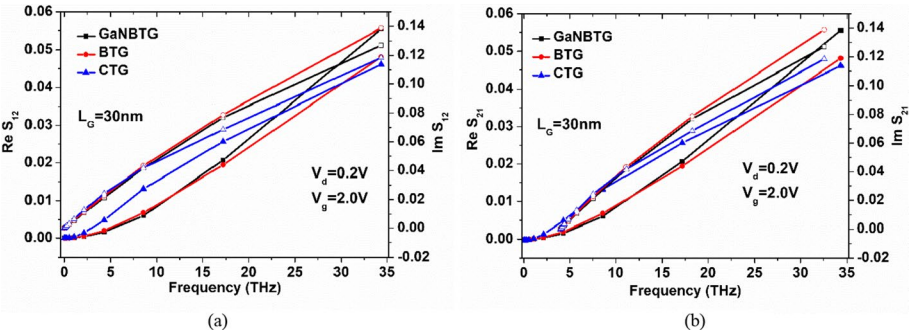


Fig. 6 **a** S_{12} and **b** S_{21} for all three devices along the frequency

3.2 S-Parameters (Scattering)

Extraction of S-parameters is the most effective technique for analysing the MOSFET's small-signal behaviour at high frequencies. The S-parameters are depicted in Figs. 5 and 6

after being modelled while maintaining the identical bias conditions for Devices 1, 2, and 3.

The input reflection coefficient at port-1 (S_{11}) and the output reflection coefficient at port-2 (S_{22}), both S-parameters, are indicators of the match's quality. S_{11} and S_{22} diminish the frequency and the reduction is greater, according to Fig. 5a and b (Eqs. 9, 12). The measure of feedback from the output of an amplifier [27] to the input of an amplifier is described by the reverse isolation parameter S_{12} and the forward transmission coefficient S_{21}. Comparing Device 1 to the other two devices, S_{12} drops [7] as shown in Fig. 6a. The decrease in S_{12} is shown as a result of the lower Z-parameter values. Simillar trend is observed in Fig. 6b for S_{21}.

$$S_{11} = \frac{(Z_{11} - Z_{01}^*)(Z_{22} + Z_{02}) - Z_{12}Z_{21}}{(Z_{11} + Z_{01})(Z_{22} + Z_{02}) - Z_{12}Z_{21}} \left(\frac{g_{ds} + w^2\beta + jw(C_{gd} + C_{sd})}{\chi} - Z_{01}^* \right) \left(\frac{w^2(\alpha + \beta) + jw(C_{gd} + C_{gs})}{\chi} + Z_{02} \right) \tag{9}$$

$$S_{11} = \frac{\left(\frac{g_{ds} + w^2\beta + jw(C_{gd} + C_{sd})}{\chi} - Z_{01}^* \right)\left(\frac{w^2(\alpha + \beta) + jw(C_{gd} + C_{gs})}{\chi} + Z_{02} \right) \left(\frac{-w\beta + jwC_{gd}}{\chi} \right)\left(\frac{w^2\beta + jwC_{gd} - g_m(1+\tau)}{\chi} \right)}{\left(\frac{g_{ds} + w^2\beta + jw(C_{gd} + C_{sd})}{\chi} + Z_{01} \right)\left(\frac{w^2(\alpha + \beta) + jw(C_{gd} + C_{gs})}{\chi} + Z_{02} \right) - \left(\frac{-w\beta + jwC_{gd}}{\chi} \right)\left(\frac{w^2\beta + jwC_{gd} - g_m(1+\tau)}{\chi} \right)} \tag{10}$$

$$S_{11} = \frac{\left(\frac{g_{ds} + w^2\beta + jw(C_{gd} + C_{sd})}{\chi} - Z_{01}^* \right)\left(\frac{w^2(\alpha + \beta) + jw(C_{gd} + C_{gs})}{\chi} + Z_{02} \right) - \left(\frac{w\beta + jwC_{gd}}{\chi} \right)\left(\frac{w^2\beta + jwC_{gd} - g_m(1+\tau)}{\chi} \right)}{\xi} \tag{11}$$

$$S_{22} = \frac{(Z_{11} + Z_{01})(Z_{22} - Z_{02}^*) - Z_{12}Z_{21}}{(Z_{11} + Z_{01})(Z_{22} + Z_{02}) - Z_{12}Z_{21}} \tag{12}$$

$$S_{22} = \frac{\left(\frac{g_{ds} + w^2\beta + jw(C_{gd} + C_{sd})}{\chi} - Z_{01}^* \right)\left(\frac{w^2(\alpha + \beta) + jw(C_{gd} + C_{gs})}{\chi} + Z_{02} \right) - \left(\frac{w\beta + jwC_{gd}}{\chi} \right)\left(\frac{w^2\beta + jwC_{gd} - g_m(1+\tau)}{\chi} \right)}{\left(\frac{g_{ds} + w^2\beta + jw(C_{gd} + C_{sd})}{\chi} + Z_{01} \right)\left(\frac{w^2(\alpha + \beta) + jw(C_{gd} + C_{gs})}{\chi} + Z_{02} \right) - \left(\frac{w\beta + jwC_{gd}}{\chi} \right)\left(\frac{w^2\beta + jwC_{gd} - g_m(1+\tau)}{\chi} \right)} \tag{13}$$

$$\left(\frac{g_{ds}+w^2\beta+jw(C_{gd}+C_{sd})}{\chi}-Z_{01}^{*}\right)\left(\frac{w^2(\alpha+\beta)+jw(C_{gd}+C_{gs})}{\chi}+Z_{02}\right)$$

$$S_{22}=\frac{-\left(\frac{w\beta+jwC_{gd}}{\chi}\right)\left(\frac{w^2\beta+jwC_{gd}-g_m(1+\tau)}{\chi}\right)}{\xi} \tag{14}$$

Thus, S_{12} & S_{21}-

$$S_{12}=\frac{2Z_{12}(R_{01}R_{02})^{1/2}}{(Z_{11}+Z_{01})(Z_{22}+Z_{02})-Z_{12}Z_{21}} \tag{15}$$

$$S_{12}=\frac{2\left(\frac{w\beta+jwC_{gd}}{\chi}\right)(R_{01}R_{02})^{1/2}}{\left(\frac{g_{ds}+w^2\beta+jw(C_{gd}+C_{sd})}{\chi}+Z_{01}\right)\left(\frac{w^2(\alpha+\beta)+jw(C_{gd}+C_{gs})}{\chi}+Z_{02}\right)}$$

$$-\left(\frac{w\beta+jwC_{gd}}{\chi}\right)\left(\frac{w^2\beta+jwC_{gd}-g_m(1+\tau)}{\chi}\right) \tag{16}$$

$$S_{12}=\frac{2\left(\frac{w\beta+jwC_{gd}}{\chi}\right)(R_{01}R_{02})^{1/2}}{\xi} \tag{17}$$

$$S_{21}=\frac{2Z_{21}(R_{01}R_{02})^{1/2}}{(Z_{11}+Z_{01})(Z_{22}+Z_{02})-Z_{12}Z_{21}} \tag{18}$$

$$S_{21}=\frac{2\left(\frac{w^2\beta+jwC_{gd}-g_m(1+\tau)}{\chi}\right)(R_{01}R_{02})^{1/2}}{\left(\frac{g_{ds}+w^2\beta+jw(C_{gd}+C_{sd})}{\chi}+Z_{01}\right)\left(\frac{w^2(\alpha+\beta)+jw(C_{gd}+C_{gs})}{\chi}+Z_{02}\right)}$$

$$-\left(\frac{w\beta+jwC_{gd}}{\chi}\right)\left(\frac{w^2\beta+jwC_{gd}-g_m(1+\tau)}{\chi}\right) \tag{19}$$

where,

$$\xi=\left(\frac{g_{ds}+w^2\beta+jw(C_{gd}+C_{sd})}{\chi}+Z_{01}\right)\left(\frac{w^2(\alpha+\beta)+jw(C_{gd}+C_{gs})}{\chi}+Z_{02}\right)$$

$$-\left(\frac{w\beta+jwC_{gd}}{\chi}\right)\left(\frac{w^2\beta+jwC_{gd}-g_m(1+\tau)}{\chi}\right)$$

and,

$$\chi = jw(C_{gs}g_{ds} + C_{gd}g_{ds} + C_{gd}g_m + \tau C_{gd}g_m) + jw^3(\alpha C_{gd} + \beta C_{gs} + C_{sd}(\alpha + \beta)) + \alpha\beta w^4 + w^2(\alpha g_{ds} + \beta g_{ds} - C_{gs}C_{sd} + \beta g_m(1 + \tau) + C_{gd}^2)$$

4 Conclusion

This research of three devices—Device 3, Device 2, and Device 1—was conducted to examine all potential performance enhancements. Tony plots are used to compare FOMs for all the devices, and FOMs for Device 1 have been compared with FOMs for Devices 3 and 2. It has been acknowledged that using GaN as a starting material rather than Si has produced observable outcomes and lessened the effects of short channels. Along with the study of Device 1's small-signal modelling, this work also covers the wave parameters, or the admittance and scattering parameters. Gain is improved by cutting back on SCEs and high ON current. Additionally, the THz range simulation results and calculated model results for small-signal parameters are achieved, validating the small-signal model. The findings thus demonstrate that Device 1 is the best choice for wireless applications at microwave frequencies.

Acknowledgements The authors are thankful to JIIT for the necessary facilities to conduct this research work.

Author Contributions AK contributed to methodology, conceptualization, and writing—original draft preparation; MMT and AK were involved in writing—review and editing.

Funding Not Applicable.

Data Availability Data will be made available on request.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

1. Arora, N. (1993). *MOSFET Models for VLSI Circuit Simulation Theory and Practice,* Springer.
2. Veeraraghavan, S., & Fossum, J. G. (1989). Short-channel effects in SOI MOSFETs. *IEEE Transactions on Electron Devices, 36*(3), 522–528.
3. Pal, A., & Sarkar, A. (2014). Analytical study of Dual Material Surrounding Gate MOSFET to suppress short-channel effects (SCEs). *Engineering Science and Technology, an International Journal, 17*(4), 205–212.
4. Kumar, A., Tiwari, B., Singh, S., Tripathi, M. M., & Chaujar, R. (2018). Radiation analysis of N-channel TGRC-MOSFET: An X-ray dosimeter. *IEEE Transactions on Electron Devices, 65*(11), 5014–5020.
5. Jo, H., Choi, S., Rhee, S., & Park, Y. J. (2017). An analytical model for the threshold voltage of intrinsic channel MOSFET having bulk trap charges. *IEEE Transactions on Electron Devices, 64*(5), 2113–2120.
6. Kumar, A. (2017). Effect of trench depth and gate length shrinking assessment on the analog and linearity performance of TGRC-MOSFET. *Superlattices and Microstructures, 109*, 626–640.
7. Kumar, A. (2020). Palladium-based trench gate MOSFET for highly sensitive hydrogen gas sensor. *Materials Science in Semiconductor Processing, 120*, 105274.

8. Kumar, A., Gupta, N., Tripathi, M., & Chaujar, R. (2019). RF noise modeling of Black Phosphorus Junctionless Trench MOSFET in strong inversion region. *Superlattices and Microstructures, 125*, 72–79.
9. Kumar, A., Tripathi, P. M., Soni, H., Chaujar, R. (2020). *Numerical Simulation and Parametric Assessment of GaN Buffered Trench Gate MOSFET for Low Power Applications*, IET Circuits, Devices Systems.
10. Ture, E., Brückner, P., Godejohann, B.-J., Aidam, R., Alsharef, M., Granzner, R., Schwierz, F., Quay, R., & Ambacher, O. (2015). High-current submicrometer tri-gate GaN high-electron mobility transistors with binary and quaternary barriers. *IEEE Journal of the Electron Devices Society, 4*(1), 1–6.
11. Kumar, A., Tripathi, S. K., Gupta, N., Tripathi, P. M., Chaujar, R. (2019). GaN Silicon-on-Insulator (SOI) N-Channel FinFET for High-Performance Low Power Applications. In *2019 IEEE 14th Nanotechnology Materials and Devices Conference (NMDC)*, IEEE, pp. 1-4
12. Pengelly, R. S., Wood, S. M., Milligan, J. W., Sheppard, S. T., & Pribble, W. L. (2012). A review of GaN on SiC high electron-mobility power transistors and MMICs. *IEEE Transactions on Microwave Theory and Techniques, 60*(6), 1764–1783.
13. Kumar, A., Gupta, N., Tripathi, S. K., Tripathi, M., & Chaujar, R. (2019). Performance evaluation of linearity and intermodulation distortion of nanoscale GaN-SOI FinFET for RFIC design. *AEU-International Journal of Electronics and Communications, 115*, 153052.
14. Adhikari, M., & Singh, Y. J. E. L. (2015). High-performance dual-channel InGaAs MOSFET for small signal RF applications. *Electronic Letters, 51*(15), 1203–1205.
15. Tsividis, Y., & McAndrew, C. (2011). *Operation and Modeling of the MOS Transistor*. Oxford Univ.
16. Silvaco, I. (2011). ATLAS User's Manual, Santa Clara, CA, Ver 5.
17. Adhikari, M., & Singh, Y. (2015). High-performance dual-channel InGaAs MOSFET for small signal RF applications. *Electronics Letters, 51*(15), 1203–1205.
18. Chalkiadaki, M.-A., & Enz, C. C. (2015). RF small-signal and noise modeling including parameter extraction of nanoscale MOSFET from weak to strong inversion. *IEEE Transactions on Microwave Theory and Techniques, 63*(7), 2173–2184.
19. Deschrijver, D., Avolio, G., Schreurs, D., Dhaene, T., Crupi, G., & Knockaert, L. (2011). Microwave small-signal modelling of FinFETs using multi-parameter rational fitting method. *Electronics Letters, 47*(19), 1084.
20. Lu, J., Wang, Y., Ma, L., & Yu, Z. (2008). A new small-signal modeling and extraction method in AlGaN/GaN HEMTs. *Solid-State Electronics, 52*(1), 115–120.
21. Roy, A. S., Enz, C. C., & Sallese, J.-M. (2006). Compact modeling of anomalous high-frequency behavior of MOSFET's small-signal NQS parameters in presence of velocity saturation. *IEEE Transactions on Electron Devices, 53*(9), 2044–2050.
22. Taher, H., Schreurs, D., & Nauwelaers, B. (2006). Extraction of small-signal equivalent circuit model parameters for Si/SiGe HBT using S-parameters measurements and one geometrical information. *AEU-International Journal of Electronics and Communications, 60*(8), 567–572.
23. Teng, M., Yue, H., Chi, C., & Xiaohua, M. (2010). A new small-signal model for asymmetrical AlGaN/GaN HEMTs. *Journal of Semiconductors, 31*(6), 064002.
24. Kumar, A., Tripathi, M., & Chaujar, R. (2018). In2O5Sn based transparent gate recessed channel MOSFET: RF small-signal model for microwave applications. *Journal of Electronics and Communications, 93*, 233–241.
25. Kumar, A., Tripathi, M.M., & Chaujar, R. (2017). Small-signal modeling of In2O5Sn based transparent gate recessed channel MOSFET for microwave/RF applications. In *Proceedings of 2017 Asia Pacific Microwave Conference*, 976–979.
26. Gupta, N., & Chaujar, R. (2016). Influence of gate metal engineering on small-signal and noise behaviour of silicon nanowire MOSFET for low-noise amplifiers. *Applied Physics A, 122*(8), 1–9.
27. Sarkar, A., & Jana, R. (2014). The influence of gate underlap on analog and RF performance of III–V heterostructure double gate MOSFET. *Superlattices and Microstructures, 73*, 256–267.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

M. M. Tripathi received the B.Tech. degree from the Madan Mohan Malaviya Engineering College, Gorakhpur, India, in 1994, and the Ph.D. degree from Gautam Buddh Technical University, Lucknow, India, in 2010. He is currently a Professor with the Electrical Engineering Department, Delhi Technological University, New Delhi, India.

Ajay Kumar received the B.Tech. degree in Electronics and Communication from UPTU, Lucknow, India in 2009 and M.Tech. degree in Nanoscience and Technology from Delhi Technological University, Delhi, India in 2014. He is currently Assistant Professor in Electronics and Communication Engineering Department in Jaypee Institute of Information Technology (JIIT), Noida and pursuing his Ph.D. degree in Electrical Engineering Department, Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India. Ajay Kumar is the Senior member of IEEE, OSA and other professional societies.

2023 3rd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 23-25, 2023

Solar Power Prediction Based on Recurrent Neural Networks Using LSTM and Dense Layer With ReLU Activation Function

Deepanshu Gupta
Department of Electrical Engineering
Delhi Technological University
New Delhi,India
deepanshugupta_2k19ee089@dtu.ac.in

Vanjari Venkata Ramana
Department of Electrical Engineering
Delhi Technological University
New Delhi,India
venkat.vr90@gmail.com

2023 3rd International Conference on Intelligent Technologies (CONIT) | 979-8-3503-3860-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CONIT59222.2023.10205605

***Abstract*— In the past decades, power production from renewable energy sources has been increasing at a tremendous rate. Such increased production had led to various benefits such as improvement of environmental conditions, production of energy independent of fossil fuels and reduction in the cost of energy production. To enjoy the benefits of renewable energy and its production in an optimum manner, it is important for us to accurately predict renewable energy production. In this paper, a model that uses deep neural network to predict solar power for two different horizons is proposed. The proposed method predicts solar power for five minutes and one hour ahead based on the observations made in the past two hours. The proposed model is executed in python software using the deep neural networks technique and is compared with an existing method in literature.**

Keywords*— *Deep neural networks, Solar power forecasting.

I. INTRODUCTION

From the last few years, there had been an increase in the demand of usage of energy which had led to greater production of non-renewable sources of energy leading to the release of harmful agents in the atmosphere such as carbon dioxide. Due to the greater usage of non-renewable sources of energy, these energy sources had been depleting day by day leading to greater production of energy from renewable sources of energy having dual benefits. One benefit is via production of energy from renewable sources there is less harm caused to the environment and second is there is less load on production of energy from non -renewable source of energy leading to preservation of non-renewable sources of energy. In order to avail the benefits of renewable energy, it is important to forecast renewable energy production. Soft computing techniques had been successful in accurately forecasting solar energy production [1].

In the current years, numerous methods of solar power prediction had been implemented [1-2]. Solar power prediction depends a lot on the physical, meteorological and climatic conditions along with the methods used for solar power prediction [3]-[5]. There are numerous methods containing neural networks that had been used in the past for solar power prediction [6]. Deep learning models are gaining popularity in various fields for making predictions using the dataset given to the model. Deep learning models had been implemented along with statistical methods for accurately predicting solar power [7]. Deep neural networks consists of artificial neural networks with two or more hidden layers to be used for solar power prediction [8].

There are various types of solar power forecasting depending upon the time horizons. Solar power prediction can be divided as very short term, short term, medium term and long term predictions [9]. There are various other ways of distinction for the prediction based on the type of method used for prediction. In [10], along with neural networks, the authors have used support vector machine (SVM) for predicting solar power. A lot of work had been done using the machine learning techniques like artificial neural networks (ANN) [11], long short term memory (LSTM), backpropagation, dense layer and many activation functions such as rectified linear unit (ReLU). In [3] authors had used ANN along with backpropagation for solar power prediction for short term.

Another use of ANN, K-nearest neighbour (KNN), support vector machine (SVM) and linear regression (LR) had been done in [12]. In [13] authors had used LR, SVM, radial basis function (RBF), multilayer perceptron (MLP), pace regression (PR), simple linear regression (SLR), least median square (LMS), additive regression (AR) for prediction of solar power. In [14] authors had used extreme machine learning (ELM), teaching learning-based optimization (TLBO), ANN for prediction of solar power for a short term duration for Chhattisgarh state.

For the task of solar power prediction, the predictions for the power output are to be made based on past observations. Since the observed data is likely to contain observations for extremely large time durations, standard neural networks are unsuitable. This is as a standard network expects a fixed dimensional input at a time, which provides the only context available from the entire series of observations. This renders such models incapable of maintaining relevant context and long-term dependencies which exist across the entire set of observations and utilize the same for future predictions.

Owing to the above problem with standard neural networks, recurrent neural networks (RNNs) have been proposed in literature [8]. In such networks, an internal context is also maintained and propagated alongside the inputs, and the model derives inference based on both the previously stored context and the current input. Further, the use of an internal context allows such networks to handle arbitrarily large sequences of data such as time series by passing individual observations for each time step separately. This makes them suitable for the task of solar power prediction, where the data is of the form of a time series.

Deep neural networks and recurrent networks have been useful in training the model and minimizing the error. Due to

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:33 UTC from IEEE Xplore. Restrictions apply.

this beneficial use of the methods, we will be using both deep neural networks and recurrent networks for solar power prediction, reducing the error making our prediction more accurate and efficient.

In this paper, deep neural networks and recurrent neural networks have been utilized to predict generated solar power ahead in future. This is beneficial for the efficient use of solar energy received by the earth as a form of renewable energy. The model targets predictions for 5-minute and 1-hour intervals based on short-term past observations across the time span of the last 2 hours.

The paper is organized as follows: Section II discusses the proposed architecture of the network used for prediction. Section III discusses results and discussion and comparisons with existing baseline approaches. The conclusion is presented in Section IV.

II. PROPOSED ARCHITECTURE

For accurate prediction of solar power, RNNs have been used widely. In the proposed model, we will be using RNN for prediction of solar power. The RNN would be consisting of a layer of LSTM followed by a layer of dense layer. At the end, a ReLU activation function is used to generate the solar power accurately.

RNNs suffer from the problem of loss of context as the size of sequences are increased, owing to the detrimental effects of gradient explosion and vanishing. To mitigate these effects, the standard RNN unit in a RNN network are replaced by more sophisticated gated units which offer better preservation of context by use of gates. One such unit is long-short term memory (LSTM). LSTM units have the capability to learn long-term dependencies from data sequences by use of gates to preserve existing context in favor of a new context generated based on the current input at time step *t*. In an LSTM cell, the inference is made using the following equations [16]:

$$\Gamma_f^{\langle t\rangle} = \sigma\left(W_f\left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_f\right) \quad \text{-(1)}$$

$$\Gamma_u^{\langle t\rangle} = \sigma\left(W_u\left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_u\right) \quad \text{-(2)}$$

$$\tilde{c}^{\langle t\rangle} = \tanh\left(W_c\left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_c\right) \quad \text{-(3)}$$

$$c^{\langle t\rangle} = \Gamma_f^{\langle t\rangle} \circ c^{\langle t-1\rangle} + \Gamma_u^{\langle t\rangle} \circ \tilde{c}^{\langle t\rangle} \quad \text{-(4)}$$

$$\Gamma_o^{\langle t\rangle} = \sigma\left(W_r\left[a^{\langle t-1\rangle}, x^{\langle t\rangle}\right] + b_o\right) \quad \text{-(5)}$$

$$a^{\langle t\rangle} = \Gamma_o^{\langle t\rangle} \circ \tanh\left(c^{\langle t\rangle}\right) \quad \text{-(6)}$$

where,

$c^{<t-1>}, a^{<t-1>}$	previous cell state
Γ_f	forget gate: determines how much of previous context to retain
$\widetilde{c^{<t>}}$	candidate cell state context for current cell
Γ_u	update gate: determines how much of candidate context to retain
$c^{<t>}, a^{<t>}$	current cell state
Γ_o	output gate: determines how much of generated context is utilized

In the proposed model depicted by Fig. 1, the architecture utilizes an LSTM layer to learn long-term dependencies from the output power observations for accurate prediction. In the primary layer, the number of units, which specifies the dimension of the cell state vectors, is set to 32, to be able to store relevant information across multiple time steps. The LSTM layer is followed by a standard feed forward layer which summarizes the information predicted by the LSTM layer and predicts the output power. Input is passed as a sliding window of size 24 across observations taken at a time interval of 5 minutes, where the model attempts to make predictions using the last 23 observations (approximately, observations from the last 2 hours).

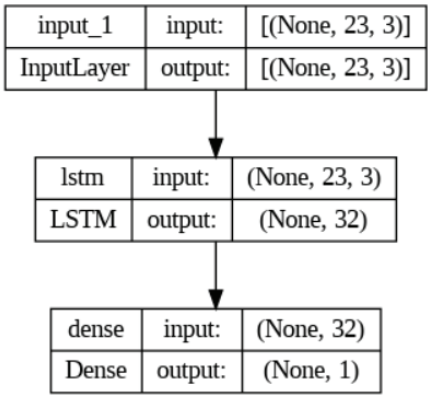


Fig. 1. The architecture of our proposed model

The model, implemented in tensorflow keras and is trained for the required number of epochs for each season over a subset of the dataset specific to each season using stochastic gradient descent with momentum, with the objective of minimizing prediction mean squared error. 0.001 is the initial learning rate, and the momentum weight is set to 0.9.

III. RESULTS AND DISCUSSION

In this section, the proposed model is compared with [15]. The data set is trained for summer, rainy and winter season. In all the power graphs along Y-axis, power is normalized between 0 to 1.

The results for the summer season are as follows:-

The solar power prediction is done for 5 minutes (Fig. 2) and 1 hour (Fig. 3) horizon. Training loss (Fig. 4) and MAE (Fig. 5) for solar power prediction are depicted by the graph below.

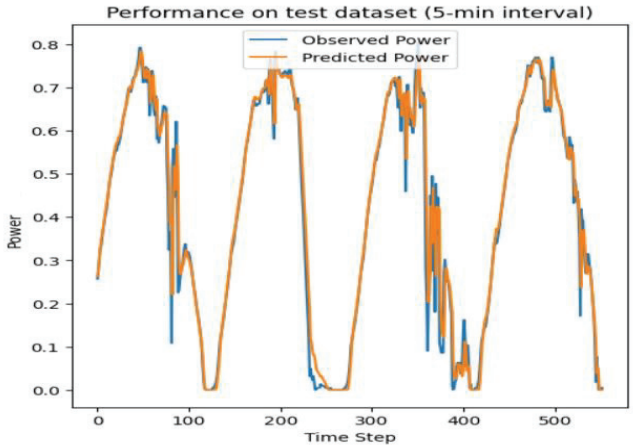


Fig. 2. Prediction of solar power for 5 minutes horizon with power and time step in seconds.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:33 UTC from IEEE Xplore. Restrictions apply.

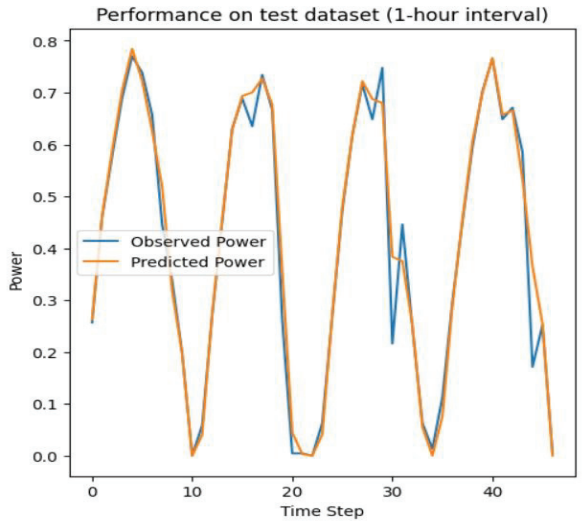


Fig. 3. Prediction of solar power for 1 hour horizon with power and time step in minutes.

Fig. 4. Training loss for 500 epochs

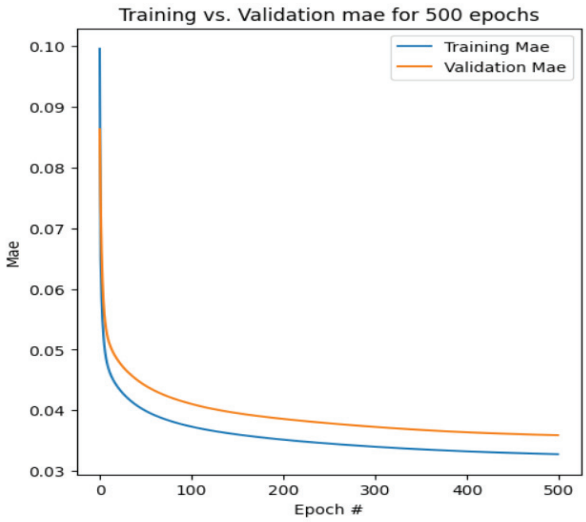


Fig. 5. Mean absolute error for 500 epochs

The MAE and RMSE for summer season is 0.02512 and 0.0472 respectively.

For the rainy season, results are as follows:

Solar power prediction is done for 5 minutes (Fig. 6) and 1 hour (Fig. 7) horizon. Training loss (Fig. 8) and MAE (Fig. 9) for solar power prediction are depicted by the graph below.

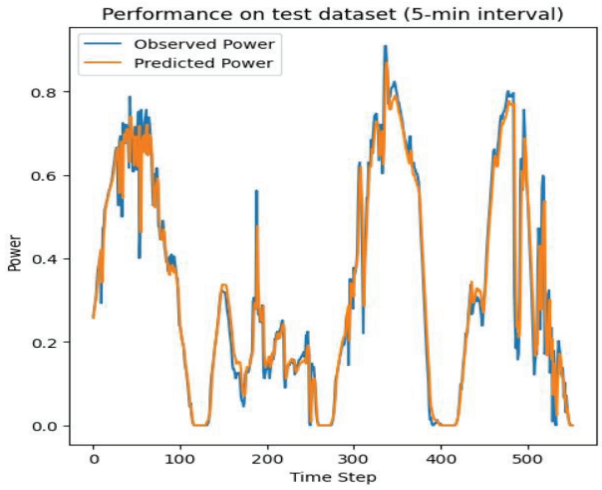


Fig. 6. Prediction of solar power for 5 minutes horizon with power and time step in seconds.

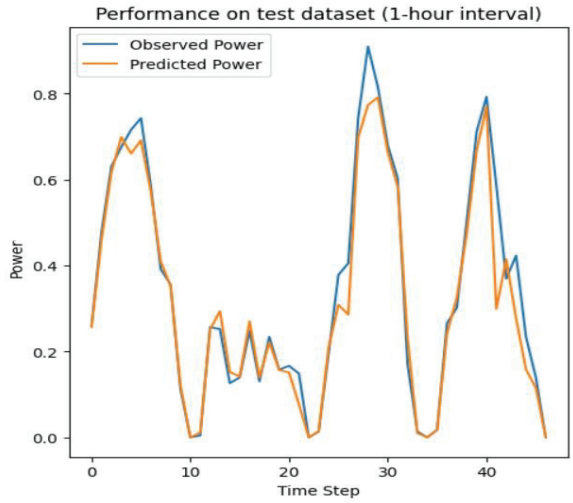


Fig. 7. Prediction of solar power for 1 hour horizon with power and time step in minutes.

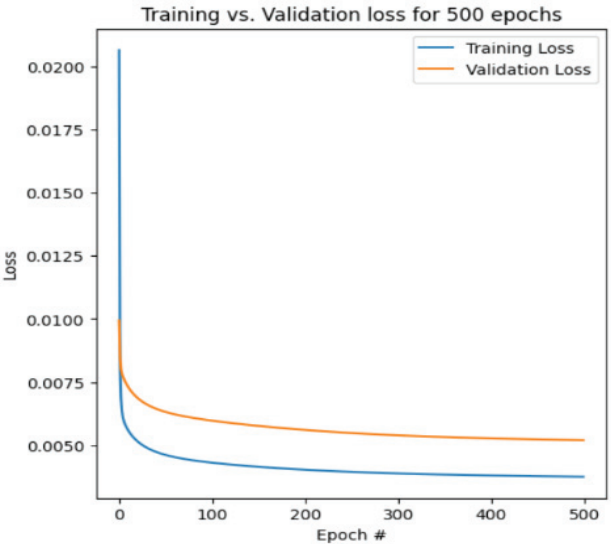


Fig. 8. Training loss for 500 epochs

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:33 UTC from IEEE Xplore. Restrictions apply.

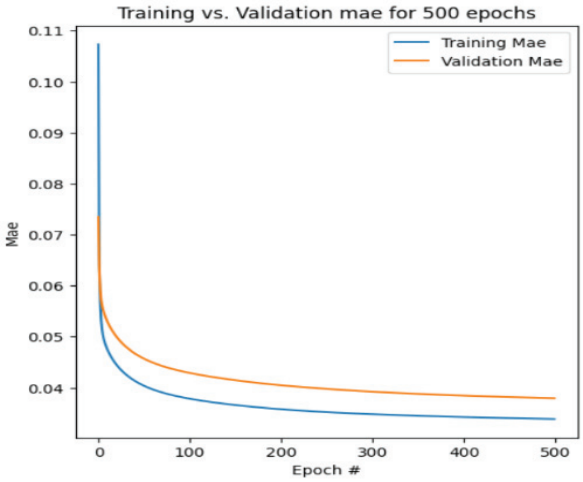


Fig. 9. Mean absolute error for 500 epochs

The MAE and RMSE for rainy season is 0.0353 and 0.0614 respectively.

For the winter season, results are as follows:-

Solar power prediction is done for 5 minutes (Fig. 10) and 1 hour (Fig. 11) horizon. Training loss (Fig. 12) and MAE (Fig. 13) for solar power prediction are depicted by the graph below.

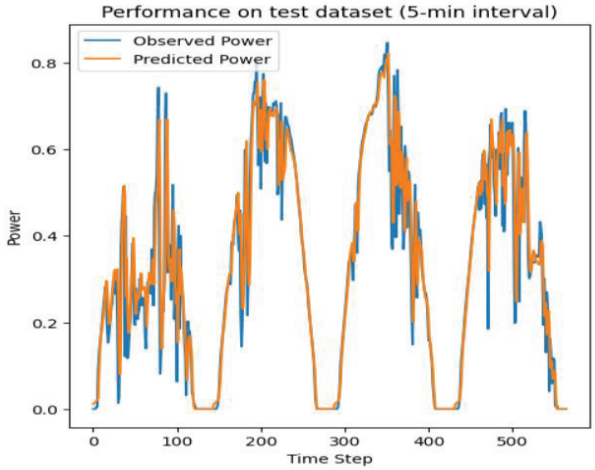


Fig. 10. Prediction of solar power for 5 minutes horizon with power and time step in seconds.

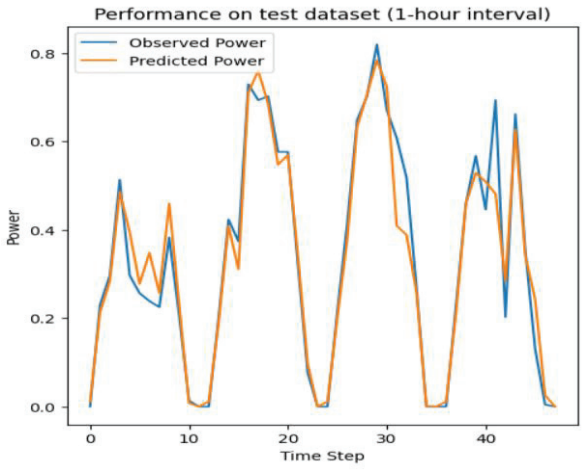


Fig. 11. Prediction of solar power for 1 hour horizon with power and time step in minutes.

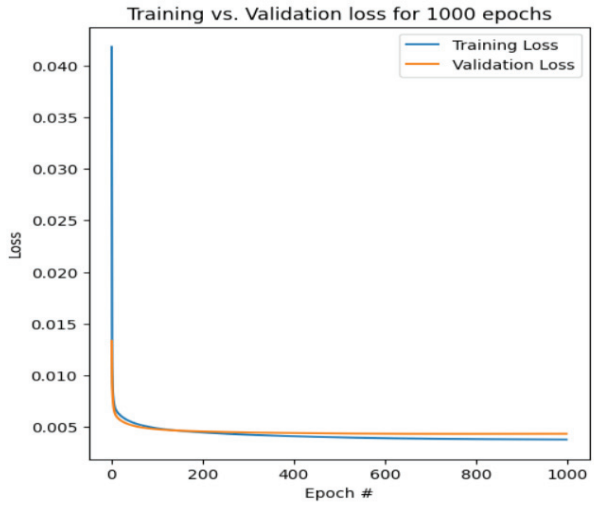


Fig. 12. Training loss for 1000 epochs

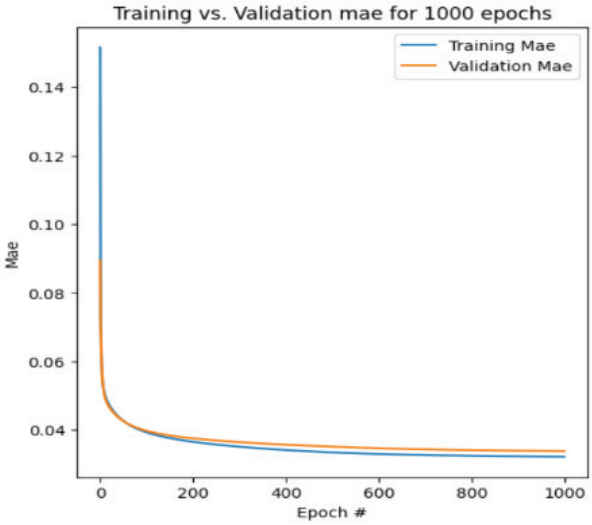


Fig. 13. Mean absolute error for 1000 epochs

The MAE and RMSE for winter season is 0.0367 and 0.0595 respectively.

IV. CONCLUSION

In this paper, deep neural networks consisting of LSTM, dense layer and ReLU activation function for prediction of solar power for two different horizons of 5 minutes and 1 hour ahead based on the observations of past 2 hours is presented. The proposed is compared with an existing method in literature and is more efficient for summer and rainy season. Results show that our proposed model is able to achieve MAE and RMSE scores of 0.02512 and 0.0472 for summer season predictions, 0.0353 and 0.0614 for rainy season predictions, and 0.0367 and 0.0595 for winter season predictions respectively. So, the proposed method is better than an existing method in literature for summer and rainy season and it is almost equivalent for the winter season.

REFERENCES

[1] Díaz–Vico, David, Alberto Torres–Barrán, Adil Omari, and Jose R. Dorronsoro. "Deep neural networks for wind and solar energy prediction." *Neural Processing Letters* 46, no. 3 (2017): 829-844.

[2] Khatib, Tamer, Azah Mohamed, Kamarulzaman Sopian, and M. Mahmoud. "Solar energy prediction for Malaysia using artificial neural networks." *International Journal of Photoenergy* 2012 (2012).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:33 UTC from IEEE Xplore. Restrictions apply.

[3] Larson, David P., Lukas Nonnenmacher, and Carlos FM Coimbra. "Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest." *Renewable Energy* 91 (2016): 11-20.

[4] Omubo-Pepple, V. B., C. Israel-Cookey, and G. I. Alaminokuma. "Effects of temperature, solar flux and relative humidity on the efficient conversion of solar energy to electricity." *European Journal of Scientific Research* 35, no. 2 (2009): 173-180.

[5] Ghosh, Santosh, Vinod Kumar Yadav, and Vivekananda Mukherjee. "Impact of environmental factors on photovoltaic performance and their mitigation strategies–A holistic review." *Renewable Energy Focus* 28 (2019): 153-172.

[6] Izgi, Ercan, Ahmet Öztopal, Bihter Yerli, Mustafa Kemal Kaymak, and Ahmet Duran Şahin. "Short–mid-term solar power prediction by using artificial neural networks." *Solar Energy* 86, no. 2 (2012): 725-733.

[7] AlKandari, Mariam, and Imtiaz Ahmad. "Solar power generation forecasting using ensemble approach based on deep learning and statistical methods." *Applied Computing and Informatics* (2020).

[8] Abuella, Mohamed, and Badrul Chowdhury. "Solar power forecasting using artificial neural networks." In *2015 North American Power Symposium (NAPS)*, pp. 1-5. IEEE, 2015.

[9] Gupta, Anunay, Anindya Gupta, Apoorv Bansal, and Madan Mohan Tripathi. "Solar Power Prediction using LTC Models."

[10] Zeng, Jianwu, and Wei Qiao. "Short-term solar power prediction using a support vector machine." *Renewable energy* 52 (2013): 118-127.

[11] Erduman, Ali. "A smart short-term solar power output prediction by artificial neural network." *Electrical Engineering* 102, no. 3 (2020): 1441-1449.

[12] Gutiérrez, Leidy, Julian Patiño, and Eduardo Duque-Grisales. "A comparison of the performance of supervised learning algorithms for solar power prediction." *Energies* 14, no. 15 (2021): 4424.

[13] Hossain, Md Rahat, Amanullah MT Oo, and ABM Shawkat Ali. "Hybrid prediction method of solar power using different computational intelligence algorithms." In *2012 22nd Australasian Universities Power Engineering Conference (AUPEC)*, pp. 1-6. IEEE, 2012.

[14] Sahu, Raj Kumar, Binod Shaw, and Jyoti Ranjan Nayak. "Short/medium term solar power forecasting of Chhattisgarh state of India using modified TLBO optimized ELM." *Engineering Science and Technology, an International Journal* 24, no. 5 (2021): 1180-1200.

[15] Mishra, Shaktinarayana, Lokanath Tripathy, Prachitara Satapathy, P. K. Dash, and Nitasha Sahani. "An efficient machine learning approach for accurate short term solar power prediction." In *2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE)*, pp. 1-6. IEEE, 2020.

[16] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9, no. 8 (1997): 1735-1780.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:21:33 UTC from IEEE Xplore. Restrictions apply.

Electrical Engineering
https://doi.org/10.1007/s00202-023-01979-0

ORIGINAL PAPER

SPV-wind-BES-based islanded electrical supply system for remote applications with power quality enhancement

Sombir Kundu[1] · Madhusudan Singh[1] · Ashutosh K. Giri[2]

Received: 2 February 2023 / Accepted: 30 July 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract
This article dealt the application of three-phase all-pass filter integrated with phase-locked loop (3Ø-APF-PLL) control approach in islanded electrical supply system for enhanced power quality (PQ). It also covers the effectiveness of proposed control under various operating conditions such as intermittent wind speed, variable solar irradiance, and unbalanced nonlinear load. An islanded electrical supply system includes wind energy generation, solar photovoltaic (SPV), and battery energy storage (BES). The proposed control technique provides a coordinated and balanced operation under a dynamical load. A 3Ø-APL-PLL-based control approach is proposed to extract the maximum value of fundamental component of load current and phase angle. The proposed control approach is utilized to operate voltage source converter (VSC) that performs multi-functions, such as reducing source current harmonics, power compensation (real & reactive), balancing of loads, and improving the overall power quality of the islanded SPV-wind-BES electrical supply system. The SPV system is developed using an incremental conductance (IC)-based maximum power point tracking (MPPT) technique to extract maximum power during the intermittent conditions of solar irradiance. A bidirectional controller (BDC) is designed to operate the charging/discharging modes of a BES. Generally, an islanded electrical supply system suffers from poor voltage & frequency regulation due to the intermittency of power generation sources, which can be solved by utilizing a BES with BDC control to manage the power leveling between the power generation sources and the unbalanced load. A 3Ø-APF-PLL control approach is compared to existing controllers like a least mean square (LMS) (Singh et al. in IET Power Electron 13:2767–2774, 2020) and second-order generalized integrator-based frequency locked loop (SOGI-FLL) (Golestan et al. in IEEE Trans Power Electron 34:213–224, 2018). Numerous parameters, such as the evaluation of the maximum value of the fundamental component, stability, steady-state error, and harmonic compensation capabilities, are taken into account when comparing 3Ø-APF-PLL to the control approaches described above. The dynamic parameters of the islanded electrical supply system are justified under the IEEE-519 standard.

Keywords All-pass filter · Renewable energy sources · Power quality · Battery energy storage

1 Introduction

The recent trend toward restructuring the global energy system to minimize emissions of carbon dioxide (CO_2) involves cooperation from populations, regions and governments, and various other stakeholders in the public and private sectors at all extents of society. Renewable energy sources (RESs) like (wind & solar) can be used to trim greenhouse gas emissions. To extract electricity from wind energy, self-excited induction generators (SEIGs) have been used comprehensively due to their low cost, lesser maintenance, and rugged structure. However, the stabilization of voltage & frequency of SEIG in stand-alone applications is a major bottleneck. A VSC-based real and reactive power compensator can progress the frequency/voltage regulation of SEIG beneath changing load conditions. The incorporation of solar PV systems at the DC side of VSC can provide real and reactive power compensation under different load demands [1]. The DC side of the VSC includes a BES to absorb the surplus power and

✉ Sombir Kundu
sskundu46@gmail.com

Madhusudan Singh
madhusudan@dce.ac.in

Ashutosh K. Giri
ashu_ee2002@rediffmail.com

[1] Delhi Technological University, Delhi, India

[2] Government Engineering College, Bharuch, Gujarat, India

Published online: 19 August 2023

thus, supply real power demands of the systems and improve the availability of the power. The charging level of the storage system can be kept to a secure extent while the VSC's voltage control capabilities are maximized with a VSI-based control framework with independent active & reactive power control [2]. A control approach based on versatile noise cancelation filters is recommended to provide a pulse signal to STATCOM. The variable frequency drive consumes real power in the over abundance of customer loads and in this way controls the frequency of the system [3]. A modified FLL-based approach is developed to enhance the synchronization and accurately evaluate the phase and frequency under unbalanced conditions. An Improved Affine Projection Versoria-based technique reduces the harmonics under an unbalanced load and enhances the quality of power in the described work [4]. A perturb & observe algorithm is implemented for the extraction of maximum power on both sides wind & solar. A combination of the wind generator (WG), diesel generator, and an SPV with BES with a minimum number of the converter is also described [5]. A fuzzy-based approach is designed to enhance the quality of power. The wind generator (WG) is associated via a step-up transformer to the network. An IC-MPPT technique-based bidirectional converter is designed to harness the peak power from an SPV. A low-voltage ride-based control is associated with a 3-phase network hybrid system to upgrade the quality of power, voltage stability, and efficiency [6, 7]. A composite sliding mode controller for a BES has been used, which supports a solar-wind electrical supply system for an isolated location [8]. An adaptive technique-based control is described to preserve the voltage of the islanded system. The control technique maintains the stator current of the isolated system under changing speeds and load perturbation three-phase four-wire (3P4W) loads [9]. For peak power point performance, a field programmable gate array-based close-loop control method is implemented by observing the power at the input terminal of the VSC [10]. A distributed electronic load controller is designed to include a suitable input filter to minimize distortion of network current and is used separately from each household to consume the excess generated power and regulate the current & voltage profile of SEIG [11].

Different types of optimization techniques have been described in the literature for the optimal use of sources in hybrid energy generation systems (HEGSs). These strategies improve the reliability of the framework and the efficacy of the isolated hybrid system [12]. A wind generator is coupled at the common point of interfacing (CPI), and the BES is associated with the bidirectional converter at the CPI. The BES is maintained the voltage & frequency profile at PCI by real/reactive power support and provides appropriate synchronization of the WG during the variation in wind speed. This control system provides a continuous power supply to loads under different operating conditions, including one voltage feedback control loop and an interfacing converter [13]. The LCL filter-based virtual resister is developed to address the problems of PQ of the wind-solar-based hybrid system [14], and control of the hybrid system in different modes is described [15]. Modeling, tuning, advantages, and disadvantages of different types of FLL have been discussed [16]. For the extraction of the load current components, a FLL-based algorithm has been developed. The FLL-based algorithm works with VSC in DSTATCOM to provide harmonic reduction, support for reactive power, operation of unit power factor, balancing of grid current, and reducing oscillations of grid current under a change in solar irradiance [17]. The system's dynamic performance is improved by DSTATCOM. By utilizing a novel harmonics isolation-based control method based on PID controller coefficients a hybrid system is developed. A differentiation-based PLL control scheme is used to regulate an islanded electrical supply system [18]. A ROGI-based, standard structure with more advanced FLLs, i.e., ROGI-FLL is also available in the literature [19]. A PLL-based voltage-controlled oscillator-less control approach is employed to extract the phase angle and fundamental component under the worst load conditions [20], and BES is developed to manage the power equilibrium between output power and the applied load [21]. A frequency adaptive control is employed to predict the hydro-generator and grid synchronization's error-free phase angles [22]. Using variable learning, an improved Adaline filter is established to improve the system's steady-state and transient response [23]. Reweighted L_1 norm penalized LMS fourth-based control is utilized to draw the maximum value of fundamental components under varying load demands [24]. An enhanced momentum LMS-based control technique-based control is described to provide a unity power factor and balance the grid current under dynamic conditions [25]. A modified multi-second-order generalized control is described for DC offset rejection [26], and the BDC control is utilized to manage the BES [27, 28]. An enhanced PLL and FLL-based technique along with VSC provide load leveling, power management, and stabilize the voltage & current profile in an isolated microgrid [29, 30]. Modeling and comparative evaluation of APF-PLL are described [31].

The main novelty of the proposed system is as follows:

- By implementing 3Ø-APF-PLL in an integrated SPV-wind-BES-based islanded electrical supply system, the faster dynamic response to regulate the voltage and frequency variations is achieved under various operating conditions of load and solar photovoltaic.
- A single-stage VSC and BDC are employed with 3Ø-APF-PLL to eliminate the PQ problems of the said system.

- The faster control of BES with VSC offers more accurate power compensation (real & reactive) at PCI, which stabilizes the frequency & voltage of an islanded electrical supply system.
- Moreover, a BDC control is implemented to maintain the charging/discharging of a BES through buck/boost modes of operation.
- An IC technique-based MPPT control is designed to vary the duty cycle in response to changing the solar irradiance.
- Moreover, a detailed comparative study and analysis of the proposed algorithm with SOGI-FLL and LMS is carried out, which will be more useful to readers.

2 System description

The parameters of the proposed three-phase three wire (3P3W) SPV-wind-BES-based islanded electrical supply system is shown in Fig. 1. The system comprises a wind turbine that works as a prime mover for SEIG, and a capacitor bank is used to provide excitation to SEIG and generate a rated voltage at no load. To maximize SPV array power during changes in solar irradiance, an IC method-based MPPT control is applied. The consumer end is coupled to a current controlled-based VSC that supplies the active/reactive power needs of SEIG and linear/nonlinear load utilizing a six-leg IGBT and interface inductors. A BES is associated at the DC end of VSC for active power storage and support during dynamic conditions. High wind speed or low load conditions cause the BES to draw excess power and charge the BES, which effectively regulates the real power and frequency of the islanded electrical supply system. Conversely, low wind speeds or high load conditions cause the BES to drain the remaining power. The maximum value of the fundamental component is extracted from the distorted and differing load current by utilizing a 3Ø-APF-PLL approach.

3 3Ø-APF-PLL control approach

The control approach that is employed to maintain voltage profile, frequency, and reduce PQ issues is crucial to the proposed SPV-wind-BES-based islanded electrical supply system's ability to operate effectively. The following summarizes the control approach:

3.1 Design of 3Ø-APF-PLL

The functional diagram of the control approach of the 3Ø-APF-PLL approach is given in Fig. 2. The proposed control is devised in two sections SRF-PLL and FFPS component detector. The FFPS component detector operates in the stationary frame ($\alpha\beta$) using the theory of instantaneous symmetrical components. According to this theory, the FFPS component of an imbalanced vector in the αβ frame can be retrieved by performing the following transformation [31].

$$\begin{bmatrix} \tilde{i}_{\alpha 1} \\ \tilde{i}_{\beta 1} \end{bmatrix} = 0.5 \begin{bmatrix} 1 & -q \\ q & 1 \end{bmatrix} \begin{bmatrix} i_{\alpha}(t) \\ i_{\beta}(t) \end{bmatrix} \tag{1}$$

where, $q = e^{-j\pi/2}$. A first-order APF is utilized to implement the 90° phase-shift operator (q) in the three-phase APF-PLL.

Taking the APF transfer function as

$$G_{\mathrm{APF}}(s) = \frac{\omega - s}{\omega + s} \tag{2}$$

Equation (1) can be expressed in space vector notation as follows:

$$\left[\tilde{i}_{\alpha 1}(s) + \tilde{i}_{\beta 1}(s)\right] = \frac{1}{2}\left[1 + j\frac{\omega - s}{\omega + s}\right]\left[i_{\alpha}(s) + j i_{\beta}(s)\right] \tag{3}$$

It exhibits unity gain with zero phase at + 50 Hz and zero gain at −50 Hz, as expected for an FFPS component detector.

3.2 Modeling and phase estimation

Suppose that the input signals of the purposed control are

$$i_{\mathrm{La}}(t) = I_m \cos(\theta_1) \tag{4}$$

$$i_{\mathrm{Lb}}(t) = I_m \cos\left(\theta_1 - {}^{2\pi}/_{3}\right) \tag{5}$$

$$i_{\mathrm{Lc}}(t) = I_m \cos\left(\theta_1 + {}^{2\pi}/_{3}\right) \tag{6}$$

where, I_m is the maximum value, and θ_1 is the phase angle of the extracted component of load current.

The 3-phase input load current signals (i_{La}, $i_{\mathrm{Lb}}$and i_{Lc}) are transformed into two-stage alpha–beta orthogonal coordinates (i_{α}, i_{β}), which are evaluated as:

$$i_{\alpha} = I_m \cos(\theta_1) \tag{7}$$

$$i_{\beta} = I_m \sin(\theta_1) \tag{8}$$

The fundamental component of computed in-phase ($I_{\alpha 1}$) and quadrature ($I_{\beta 1}$) current equations are evaluated as:

$$I_{\alpha 1} = I_m \cos(\theta_1) \tag{9}$$

$$I_{\beta 1} = I_m \sin(\theta_1) \tag{10}$$

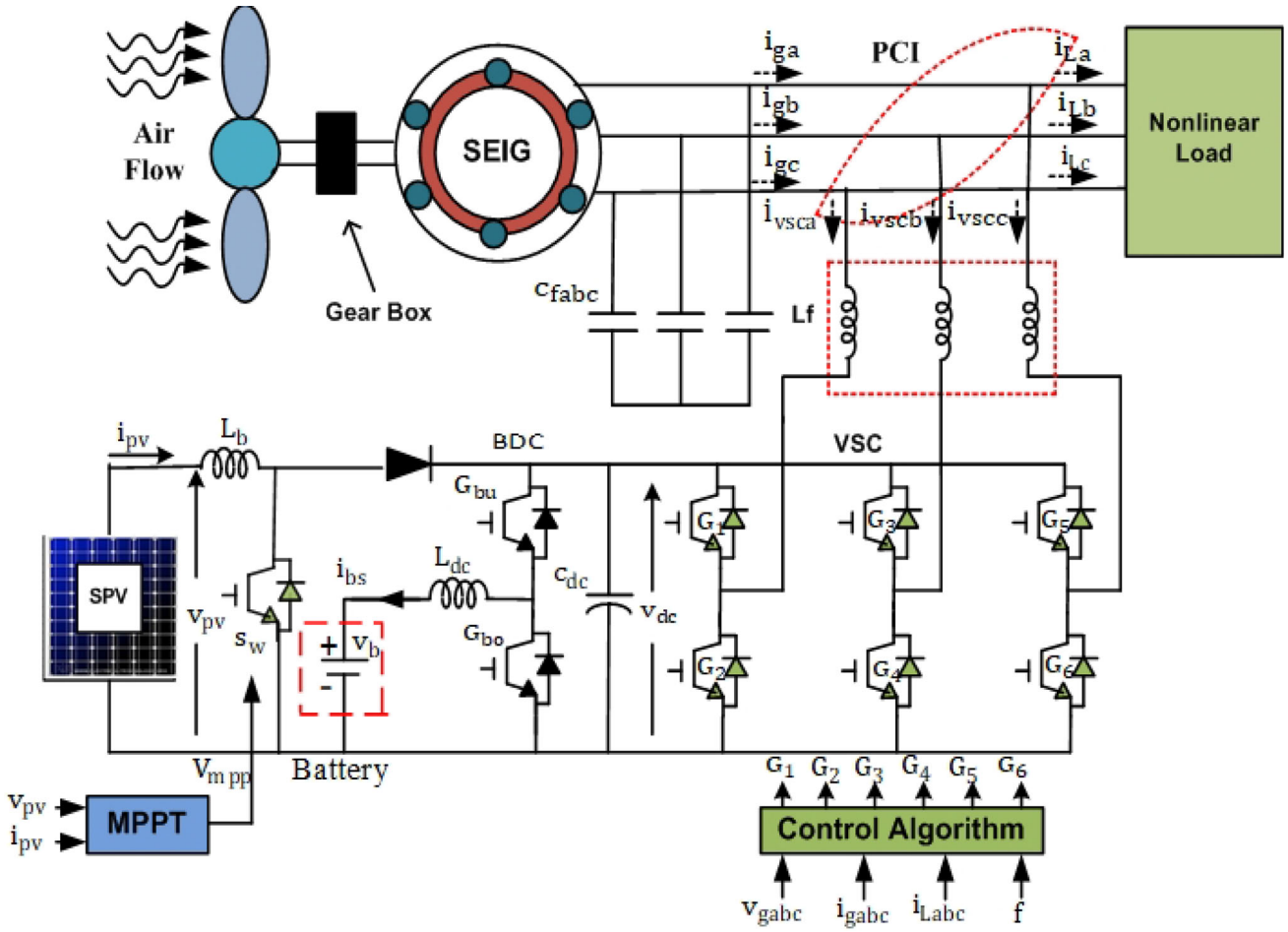


Fig. 1 Block diagram of 3Ø-APF-PLL-based islanded electrical supply system

where the maximum value of the fundamental component and phase angle of load current are evaluated as [17]

$$I_m = \sqrt{I_{\alpha 1} + I_{\beta 1}} \tag{11}$$

$$\theta_1 = \tan^{-1}(I_{\beta 1}/I_{\alpha 1}) \tag{12}$$

3.3 Computation of voltage amplitude and unit vector estimation

At PCI, the measured line voltages are transformed to phase voltage, which takes the following form:

$$v_{\mathrm{ga}} = \frac{1}{3}(2v_{ab} + v_{bc}),\quad v_{\mathrm{gb}} = \frac{1}{3}(-v_{ab} + v_{bc}),$$
$$and\ v_{\mathrm{gc}} = \frac{1}{3}(-v_{ab} - 2v_{bc}) \tag{13}$$

The amplitude of terminal voltage (V_m) can be estimated as:

$$V_m = \left[2\left(v_{\mathrm{ga}}^2 + v_{\mathrm{gb}}^2 + v_{\mathrm{gc}}^2\right)/3\right]^{\frac{1}{2}} \tag{14}$$

where, the generated voltages v_{ga}, v_{gb} and v_{gc} are 3-phase instantaneous voltage values. The voltage amplitude is estimated by dividing the instantaneous voltages (V_m), the unit vectors may be evaluated as

$$p_{\mathrm{ua}} = {}^{v_{\mathrm{ga}}}/_{V_m},\ p_{\mathrm{ub}} = {}^{v_{\mathrm{gb}}}/_{V_m},\ p_{\mathrm{uc}} = {}^{v_{\mathrm{gc}}}/_{V_m} \tag{15}$$

Utilizing the in-phase vectors, quadrature unit voltage vectors are computed and taken as follows:

$$\begin{pmatrix} q_{\mathrm{ua}} \\ q_{\mathrm{ub}} \\ q_{\mathrm{uc}} \end{pmatrix} = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 & 0 & 1 \\ \sqrt{3} & 1 & -1 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} p_{\mathrm{ua}} \\ p_{\mathrm{ub}} \\ p_{\mathrm{uc}} \end{pmatrix} \tag{16}$$

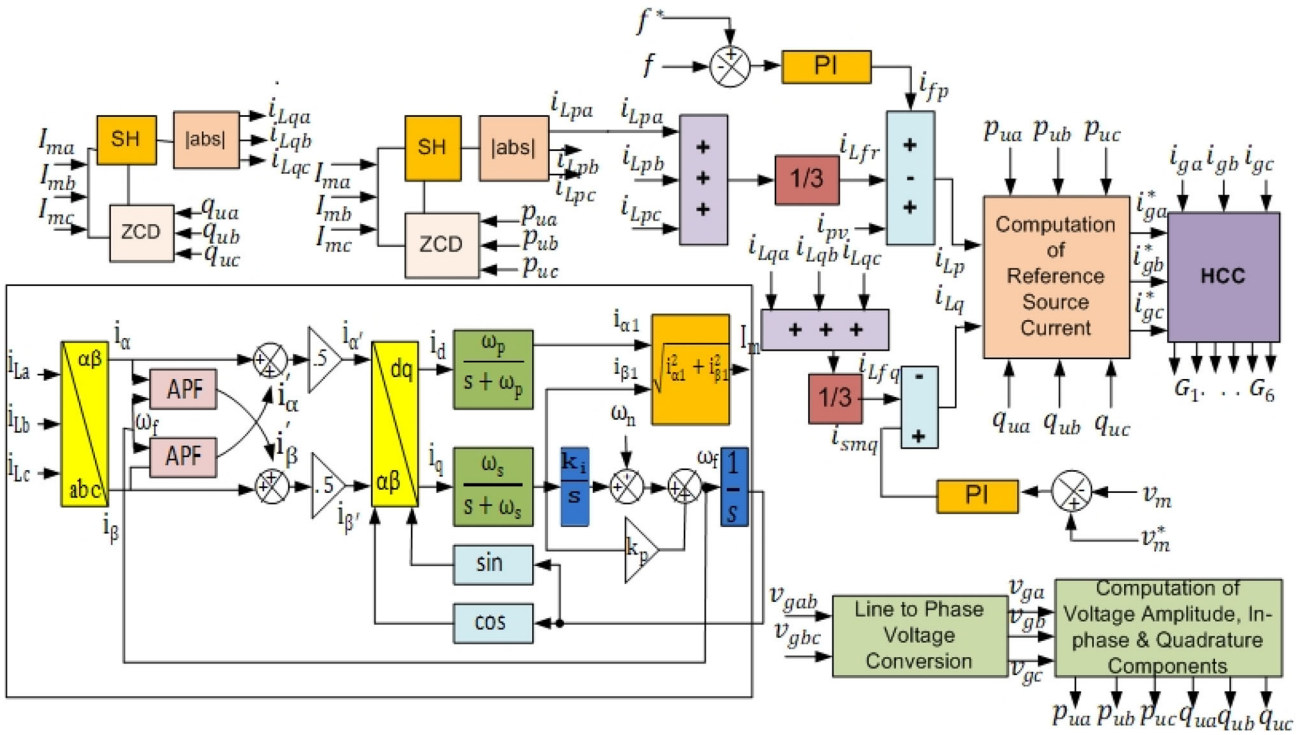


Fig. 2 Control scheme of an islanded electrical supply system

where, (p_{ua}, p_{ub}, p_{uc}) is in-phase unit voltage vectors, and (q_{ua}, q_{ub}, q_{uc}) is the quadrature phase unit voltage vectors [2].

3.4 Gate pulses generation

A 3Ø-APF-PLL-based approach is developed to extract the maximum value of the fundamental component (I_m), to obtain the instantaneous value. The fundamental load current magnitude (I_m) is multiplied with a fundamental quadrature component ($I_{\beta 1}$). The expansion of PV current (i_{pv}) and the yield of the PI controller (i_{fp}) are minus from the active current component (i_{Lfr}) of fundamental load current, to obtain the total active component of reference source current as shown in Fig. 2. The comparison between reference frequency (f*) and observed frequency (f) is processed by the PI controller.

The frequency at the sth sampling is obtained and the difference between the two frequencies is taken as:

$$f(s) = f^*(s) - f(s) \tag{17}$$

where the power loss element of current is represented by i_{fp} (s), and the PI controller gain constants are k_d and k_i. The active element of the reference source current is calculated as follows:

$$i_{\mathrm{Lp}} = i_{\mathrm{fp}} + i_{\mathrm{pv}} - i_{\mathrm{Lfr}} \tag{18}$$

$$i^*_{\mathrm{Lp}} = i_{\mathrm{Lp}} P_{\mathrm{uabc}} \tag{19}$$

The distinction between the reactive component (i_{Lfq}) of the fundamental load current, and the output current (i_{smq}) is utilized to compute the reference quadrature current (i_{Lq}). The current component (i_{smq}) computation can be derived as follows. The voltage error (v_{terr}) of a PCI voltage at sth sampling instant is computed as:

$$V_{\mathrm{merr}}(s) = V^*_m(s) - V_m(s) \tag{20}$$

The magnitude of the reference quadrature current (i_{Lq}) component is estimated as $i_{\mathrm{Lq}} = i_{\mathrm{smq}} - i_{\mathrm{Lfq}}$ and is given below:

$$i^*_{\mathrm{Lq}} = i_{\mathrm{Lq}} q_{\mathrm{uabc}} \tag{21}$$

The reference source currents are estimated using the following relationship

$$i^*_{\mathrm{gabc}} = i^*_{\mathrm{Lp}} + i^*_{\mathrm{Lq}} \tag{22}$$

For the generation of gate pulses, the sensed source currents (i_{gabc}) and extricated reference source current (i^*_{gabc}) are compared in hysteresis current controller, and gating signals (G_1, G_2, G_3, G_4, G_5, G_6) are produced to control IGBTs in VSC [20].

3.5 MPPT Control

The SPV system is designed with the IC-MPPT method to extricate maximum power from an SPV and to vary the duty cycle. The maximum power at the maximum power point is reached when $(dp/dv = 0)$. In mathematical formulation, an IC process is described as:

$$\frac{\mathrm{d}i_{\mathrm{pv}}}{\mathrm{d}v_{\mathrm{pv}}} > -\frac{i_{\mathrm{pv}}}{v_{\mathrm{pv}}} \text{ then } d_{\mathrm{new}} = d_{\mathrm{old}} + \Delta d \tag{23}$$

$$\frac{\mathrm{d}i_{\mathrm{pv}}}{\mathrm{d}v_{\mathrm{pv}}} = -\frac{i_{\mathrm{pv}}}{v_{\mathrm{pv}}} \text{ then } d_{\mathrm{new}} = d_{\mathrm{old}} \tag{24}$$

$$\frac{\mathrm{d}i_{\mathrm{pv}}}{\mathrm{d}v_{\mathrm{pv}}} < -\frac{i_{\mathrm{pv}}}{v_{\mathrm{pv}}} \text{ then } d_{\mathrm{new}} = d_{\mathrm{old}} - \Delta d \tag{25}$$

$d_{\mathrm{new}}, d_{\mathrm{old}}$ and Δd are the evaluated, old and change in duty ratio of the boost converter. The MPPT is tracked at every step by equating the conductance to an IC algorithm. MPPT's main objective is to obtain the boost converter's duty ratio to achieve maximum power from the SPV. An integral controller and proportional-integral are employed to minimize the steady-state error and to reduce signal error [27].

3.6 Bidirectional control

The BDC control is implemented to control the voltage at DC-link and control also manages the charging/discharging of the BES through buck-boost mode operation. The charging and discharging of BES are determined by the load. As a result, the BDC starts working in boost mode, and current flows from the BES to the applied load, and the BES switches to discharge mode. During reduced load conditions, the BDC operates in buck mode, and the BES is switched to charging mode. The extra power is then utilized by BES. As a result, the BDC serves two functions: regulating the BES current and regulating the DC-link voltage. The governing equation of the BDC is expressed as

$$\frac{v_{\mathrm{dc}}}{i_b} = \frac{(1-\delta)(R_{\mathrm{dc}} - sL_{\mathrm{dc}})}{1 + (sR_{\mathrm{dc}}C_{\mathrm{dc}}) + (1-\delta)^2} \tag{26}$$

$$\frac{v_{\mathrm{dc}}}{\delta_{\mathrm{dc}}} = \left(\frac{2 + sR_{\mathrm{dc}}C_{\mathrm{dc}}}{\left(s^2 L_b C_{\mathrm{dc}} + \frac{sL_{\mathrm{dc}}}{R_{\mathrm{dc}}} + (1-\delta)^2\right) + (1-\delta)^2}\right)\left(\frac{v_b}{R_{\mathrm{dc}}(1-\delta)}\right) \tag{27}$$

where C_{dc}, L_{dc}, δ_{dc} and R_{dc} are the DC-link voltage, an inductor of BDC, the duty of BDC, and output effective resistance $(R_{\mathrm{dc}} = v_{\mathrm{dc}}^2/p_{\mathrm{dc}})$. The equation of PI controller manages the DC-link voltage is given as

$$i_b^*(s) = i_b^*(s-1) + k_{\mathrm{pi}}\{v_{\mathrm{dcer}}(s) - v_{\mathrm{dcer}}(s-1)\} + k_{\mathrm{pv}} v_{\mathrm{dcer}}(s) \tag{28}$$

where

$$v_{\mathrm{dcer}}(s) = v_{\mathrm{dc}}^*(s) - v_{\mathrm{dc}}(s) \tag{29}$$

where i_b and i_b^* are the sensed BES current and reference BES current. The comparison of the saw-tooth wave and the output of PI controller is utilized to generate the BDC gating pulses [30].

4 Simulation results

Using MATLAB, the detailed simulation results for the wind, solar PV, BES, and nonlinear load-based stand-alone electrical supply system are examined in this section. In this work, the system is simulated with changes in wind speed/solar irradiance and unbalanced load. In addition, the internal parameters of the designed 3Ø-APF-PLL control strategy are thoroughly examined.

4.1 Internal parameters of the control approach

Figure 3a–c represents the internal control parameters of the 3Ø-APF-PLL-based islanded electrical supply system under an unbalanced load. The load of phase 'a' is detached at $t =$ 3.3 s. The control signals i_α, i_β, $i_{L\alpha 1}$, $i_{L\beta 1}$, i_q, and I_m are shown in Fig. 3a. These control signals represent the dynamics of various intermediate variables of the control scheme under the removal of load. Due to the elimination of load the value of i_α is reaches zero at $t = 3.3$ s. The control signals i_{smq}, p_{uabc}, q_{uabc}, i_{Lfr}, i_{Lfq} and ω_g are shown in Fig. 3b. The control signal shows the extracted maximum value of load current, AC PI controller current, and active/quadrature unit vectors. The control signals $i_{\mathrm{pf}}, i_{\mathrm{Lp}}, i_{\mathrm{Lq}}, i_{\mathrm{Lp}}^*, i_{\mathrm{Lq}}^*$, and i_{gabc}^* are shown in Fig. 3c to represent the extracted reference current is completely sinusoidal during distorted load conditions and the real/reactive reference current, DC PI controller current, and extracted reference current.

4.2 Dynamic behavior at the change in wind speed

Figure 4 depicts the dynamic behavior of an islanded electrical supply system with changes in wind speeds and constant solar irradiance feeding a linear/nonlinear load. The controller stabilizes the frequency (f) and BES voltage (v_b) or DC-link voltage under the intermittent condition of wind speed. The wind speed is suddenly changed from 14 to 11 m/s at $t = 3.3$ s. Results in a decrement in wind power generation and deficit load power, which are amplified by the BES through VSC; therefore, compensator current is slightly increased, and BES charging current is reduced after $t =$ 3.3 s. During declination in wind speed, the BES switches from charging to draining mode and maintains the balance

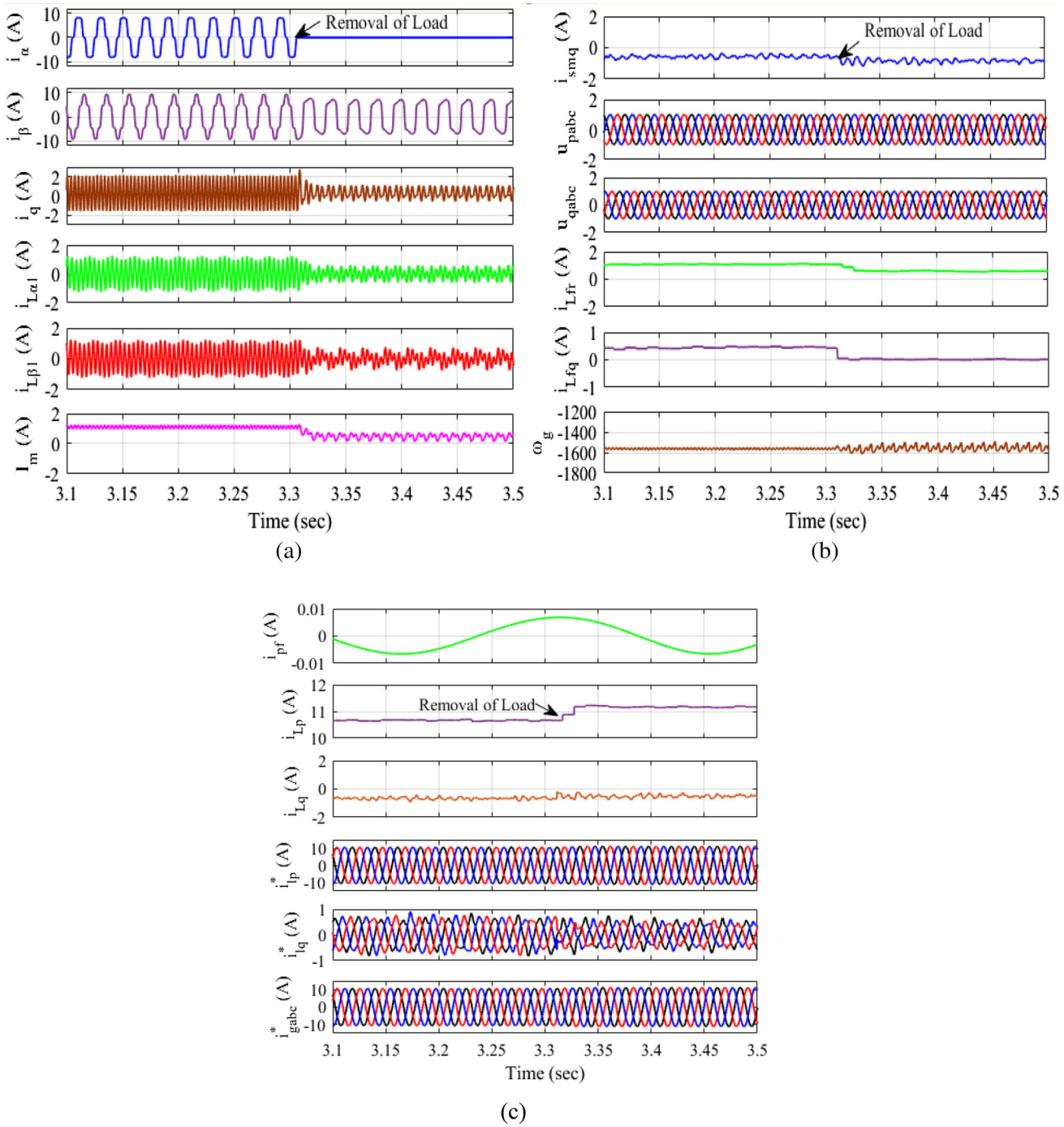


Fig. 3 **a–c** Estimation of fundamental components of the load current using 3Ø-APF-PLL control algorithm, **b** estimation of real/reactive unit vectors and AC PI controller current with other intermediate signals using 3Ø-APF-PLL **c** DC PI controller current and reference current generation using 3Ø-APF-PLL

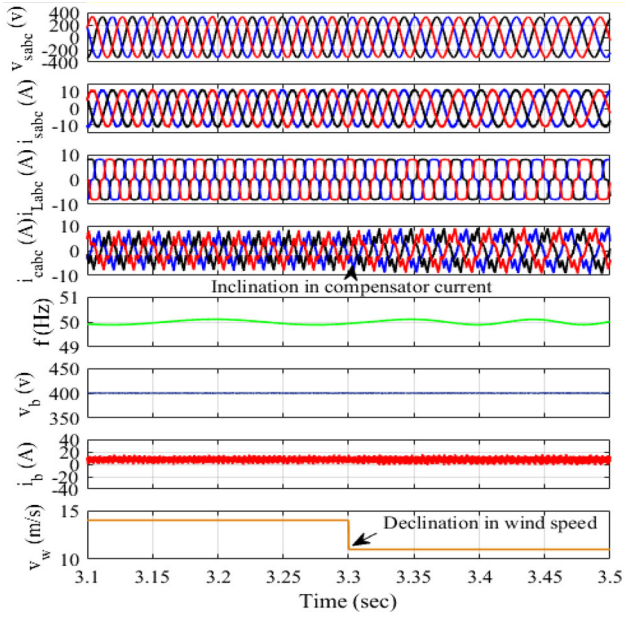
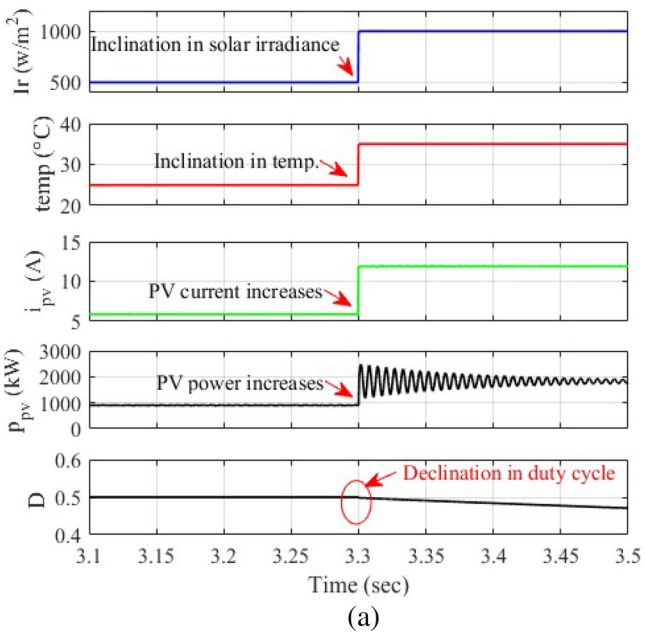


Fig. 4 Dynamic behavior at the change in wind speed

of power at PCI. A 3Ø-APF-PLL-based approach with VSC effectively maintains the propoosed system voltage & current profile during declination in wind speed.

4.3 Dynamic behavior under the change in solar irradiance

The dynamic behavior of the proposed SPV-wind-BES-based islanded electrical supply system is depicted in Figs. 5a, b with variable solar irradiations 500–1000w/m^2 and temperature 25–35 ºC, with the wind speed maintained at 14 m/s supplying a linear/nonlinear load. Figure 5a shows the parameters of the SPV system such as PV current (i_{pv}) and PV power (P), which increases and decreases in the duty cycle due to changes in solar irradiance at $t = 3.3$ s. During the changing condition of RESs (wind and solar), the BES is employed to maintain the power equilibrium between the applied load and the islanded system. The controller of the islanded electrical supply system monitors the system frequency, voltage, and harmonics, as depicted in Fig. 5b. Due to the inclination of solar power, the BES charging current, PCI current, and compensator current are gradually increased, and the controller maintains other parameters of the islanded system are kept constant. As solar power increases, the BES current shifts from draining to charging mode, and the remaining power is consumed by the BES.



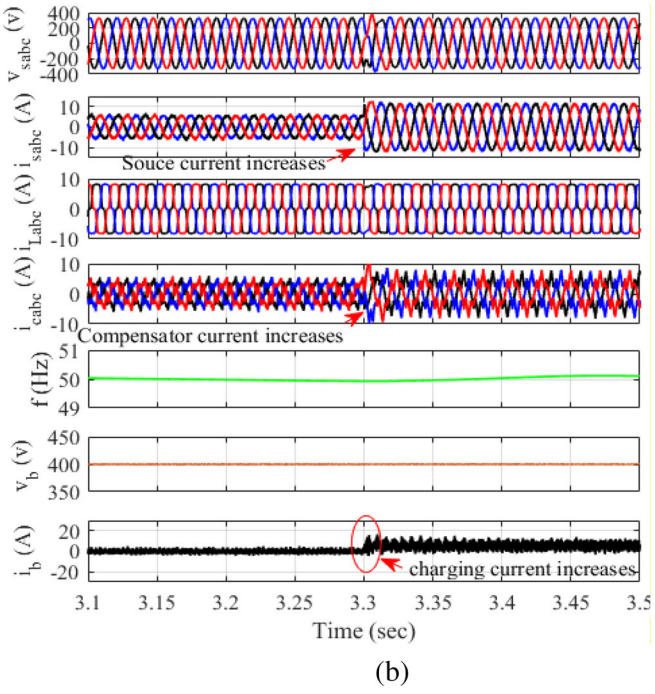


Fig. 5 **a** and **b** Dynamic behavior under the change in solar irradiance

4.4 Dynamic behavior at an unbalanced load

Figure 6 depicts the dynamic behavior of a SPV-wind-BES-based islanded electrical supply system with a constant wind speed/solar irradiance feeding an unbalanced load. The load of phase 'a' is eliminated at $t = 3.2$ s to $t = 3.4$ s. When the load is removed, the BES charging current increases and extra power is given to the BES. The load current is decreased due to disconnection of load at $t = 3.2$ s to $t = 3.4$ s. A 3Ø-APF-PLL-based controller regulates the islanded electrical supply system frequency and stabilizes the voltage profile, current profile, and DC-link voltage or BES voltage during

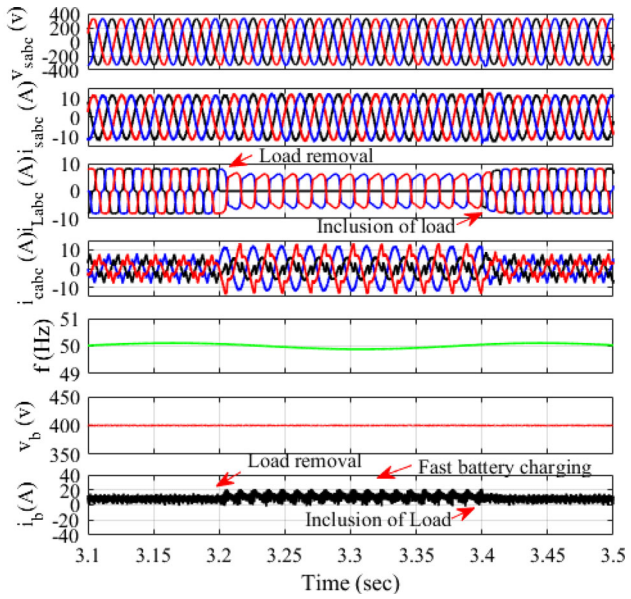


Fig. 6 Dynamic behavior at unbalanced load

the fluctuations in load. The source current stays sinusoidal and balanced even under unbalanced load situations.

4.5 Dynamic behavior at a change in RESs and load

Figure 7a, b depicts the behavior of the islanded electrical supply system under variations in energy sources feeding an unbalanced load. At t = 3.3 s, the declination in solar irradiance and temperature ranges from 1000 to 500w/m^2 and 35–25 °C. As a result, the PV current, power, and duty cycle are lowered. Figure 7b depicts the wind speed declination from 14 to 11 m/s at t = 3.3 s, as well as the imbalanced load (the burden of phase 'a' is abolished from t = 3.2 s to t = 3.4 s). During the altering condition of RESs and load, a 3Ø-APF-PLL-based controller successfully stabilizes the current & voltage profile of SEIG. Due to the decline in solar irradiance and unbalanced load, the source current of the islanded system is decreased, and the load current is also reduced at $t = 3.2$ s to $t = 3.4$ s. The BES current is gradually increasing during load removal, which means the additional power is delivered to the BES for charging. The BES current is reduced during declination in RESs at t = 3.3 s, i.e., the BES stored power is supplied at PCI to manage the level of power between the RESs and the applied load.

4.6 Harmonics investigation of islanded system

Figure 8a–c shows the harmonics investigation of PCI voltage (v_{gc}), PCI current (i_{gc}), and load current (i_{Lc}) of phase 'c' are 1.42, 4.32, and 25.06%, respectively. When the nonlinear load is applied to the 3Ø-APF-PLL-based islanded electrical supply system, the THD of PCI current is less than 5%, which satisfies the limit of the IEEE-519 standard.

4.7 Comparative analysis of 3Ø-APF-PLL, SOGI-FLL, and LMS in islanded electrical supply system

Figure 9a, b compares the performance of 3Ø-APF-PLL, SOGI-FLL, and LMS under an imbalanced load. The proposed method has the benefit of quick convergence, fast dynamic response, and fast BES charging response under varying load demands. Figure 9a depicts the comparative simulation study of the extracted fundamental load current component. Under dynamic load circumstances, the 3Ø-APF-PLL control exhibits reduced steady-state oscillations, quicker convergence, and dynamic responsiveness. The computation time of the 3Ø-APF-PLL-based control is one cycle under dynamic conditions. Figure 9b compares the response of BES charging current for 3Ø-APF-PLL, SOGI-FLL, and LMS when the load of one phase is detached from $t = 3.2$ s to $t = 3.4$ s. During the removal of load, the 3Ø-APF-PLL control technique provides fast dynamic response and convergence. Table 1 compares the response of the 3Ø-APF-PLL, SOGI-FLL, and LMS control algorithms.

5 Test results

The proposed 3Ø-APF-PLL control approach is developed on hardware setup to examine the response of the islanded system as shown in Fig. 10. The LEM LA-25-based Hall-effect sensors have been employed to observe voltages & currents in the islanded system. In the proposed 3P3W islanded system, the Hall-effect-based sensors are employed to monitor the source currents of phases 'a' and 'b,' respectively. In the described setup, three current sensors (LEM LA-25) are utilized to sense the load currents in the islanded system. The output of the current sensors is connected to the analog-to-digital converter (ADC) channel of the dSPACE-1104, through suitable buffer circuitry. The buffer circuit ensures that the voltage signal from the current sensors is properly conditioned and compatible with the ADC requirements. The buffer circuit is designed with operational amplifiers (OP-07). The buffer circuit is designed with appropriate resistance values to achieve unity gain, meaning the output voltage of the buffer is equal to the input voltage. The low output resistance of the buffer ensures that the load connected to it does not significantly affect the signal voltage. The extremely high input resistance of the buffer prevents loading effects on the current sensors, ensuring accurate current measurement. The dynamics performance of the system was examined using four different channels of the DSO-X 2024A. The outcomes were measured while the system was

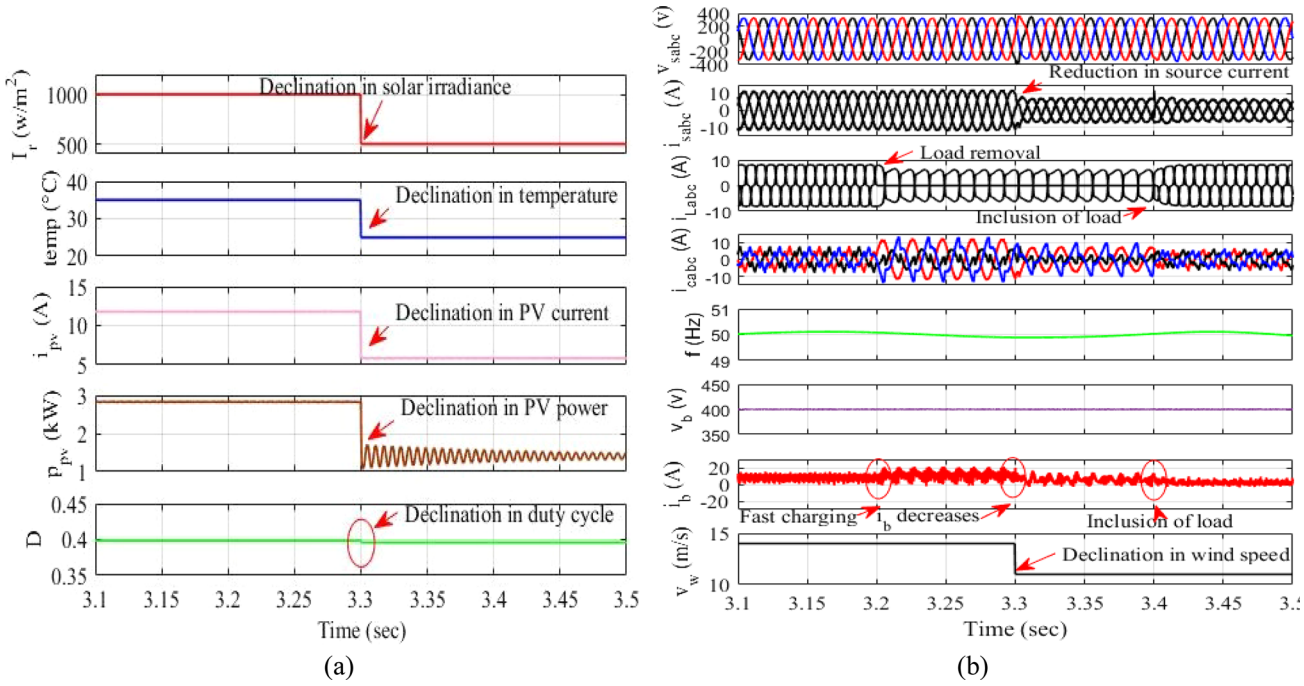


Fig. 7 **a** and **b** Dynamic behavior at the change in RESs and unbalanced load

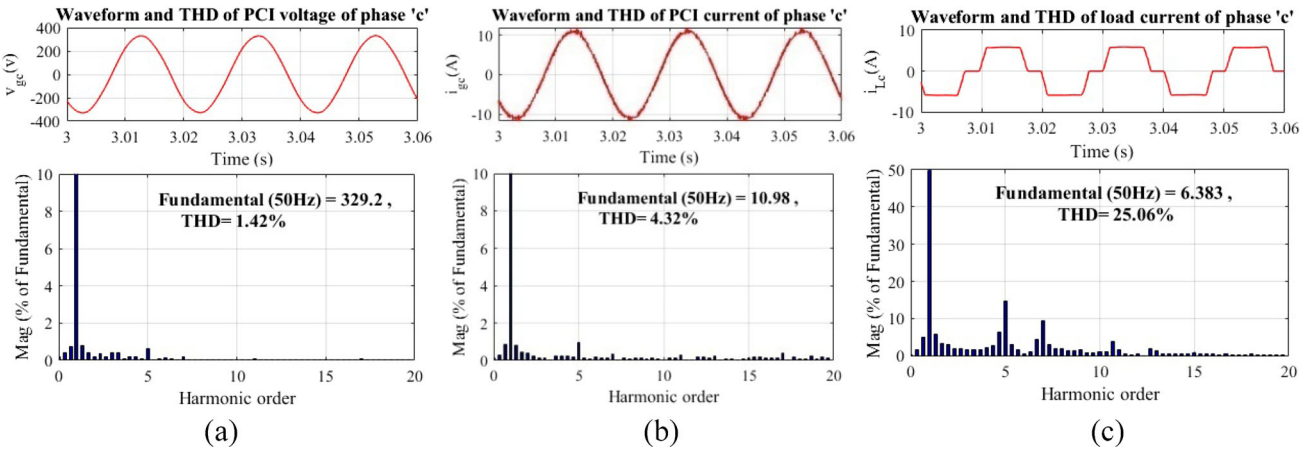


Fig. 8 **a–c** Harmonics investigation of islanded system

operating in a steady state utilizing the PQ analyzer. The examined test results and harmonic spectra of the islanded system are depicted in Figs. 11a–f and 12a, b.

5.1 Dynamic behavior of SPV-Wind-BES islanded electrical supply system

The dynamic behavior of the islanded system is observed using DSO-X 2024A under the unbalanced condition of the load. The observed results are depicted in Fig. 11a–f. Figure 11a depicts the load current of 3-phases (a, b & c) and DC-link voltage. During the insertion of load phase 'b,' the load current of all phases is increased, and the DC-link voltage is maintained constant by the proposed controller. Figure 11b depicts the DC-link voltage, PCI voltage, PCI current, and load current. Due to the sudden inclination in a load of phase 'a,' the PCI current and load current of the islanded system are increased. In Fig. 11c, during an inclination in load, the PCI current and load current are increased. A3Ø-APF-PLL-based controller with VSC maintains the PCI voltage sinusoidal, and the DC-link voltage is gradually reduced under inclination in load. Figure 11d depicts the load current and PCI current is reduced due to the removal of a load of phase 'a.' Figure 11e shows a sudden change in load, the DC-link voltage is suddenly increased and the PCI voltage is maintained sinusoidal. Figure 11f depicts the

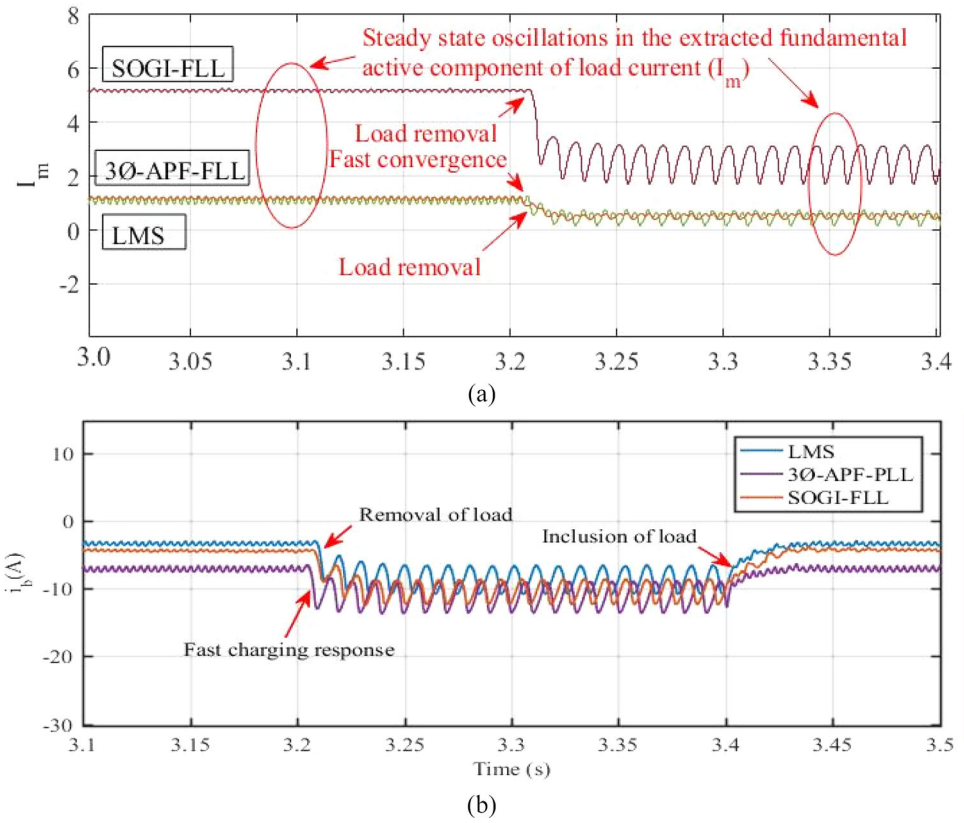


Fig. 9 **a** and **b** Comparison in the response of fundamental load current component (I_m) and (b) BES charging response for 3Ø-APF-PLL, SOGI-FLL, and LMS under disconnection of a load of phase 'a'

Table 1 Comparative performances of 3Ø-APF-PLL, SOGI-FLL, and LMS control algorithms

S.no	Parameters	3Ø-APF-PLL	SOGI-FLL	LMS
1	Computation time for fundamental extraction of load current component	One cycle	Three cycle	Three cycle
2	Steady-state oscillations	Less	Medium	High
3	Stability	Best	Good	Bad
4	Response	Fast	Medium	Medium
5	THD obtained			
	PCI voltage	1.42%	2.85%	3.73%
	PCI current	4.32%	4.89%	5.19%
	Load current	25.06%	26.18%	24.79%

parameters of the islanded system; the load of phases '*a*' and '*b*' is interfaced with the system. Due to the inclination of the two-phase load, the PCI current and load current are increased. During all circumstances, the proposed control has the versatile ability to manage all the parameters of the islanded system. Moreover, the 3Ø-APF-PLL-based control with VSC supplies improved PQ, suppression of harmonics, leveling of load, power compensation, and voltage & frequency stabilization. The BES system stored the unused power and supplies the power through VSC under varying load demands.

5.2 Harmonic spectrum

The harmonic spectrum of islanded system is analyzed using a 3-phase power analyzer. The recorded results show the THD of PCI (voltage & current), and load current are 1.84, 4.24, and 25.62%, respectively. The THD results are taken

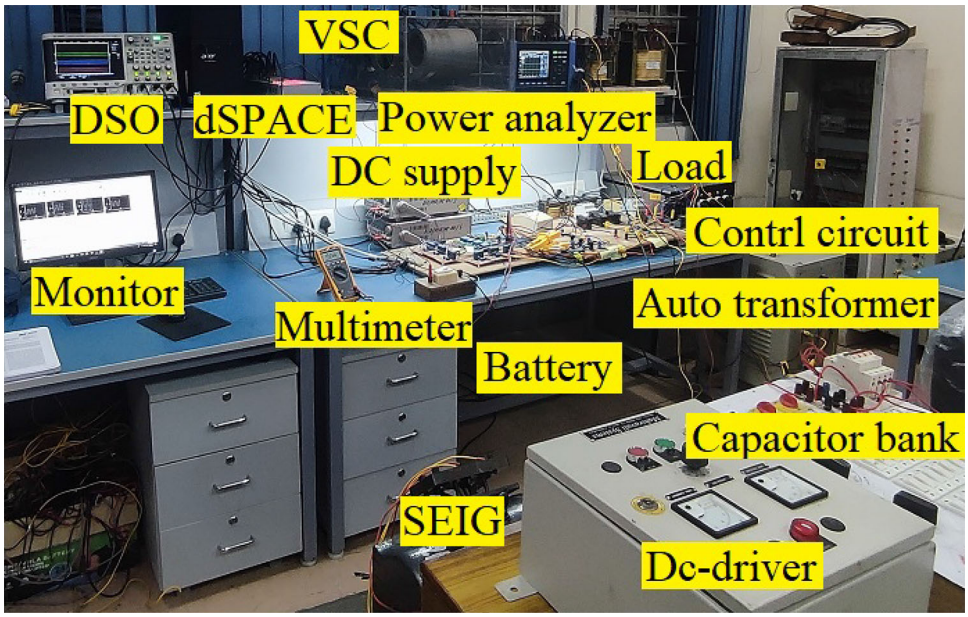


Fig. 10 Experimental prototype of the islanded electrical supply system

under a highly distorted load current. The THD of PCI current is under 5% and fulfills the criteria of the IEEE-519 standard. The 3Ø-APF-PLL-based controller with VSC compensates the harmonics injected by non-sinusoidal load current (25.62%) and maintains the PCI current sinusoidal.

6 Conclusion

This proposed islanded electrical supply system performance is analyzed with a 3Ø-APF-PLL control approach under a change in wind speed/solar irradiance feeding an unbalanced distorted load. The 3Ø-APF-PLL approach is implemented to extract the maximum value of fundamental component and estimation of active & reactive components for the generation of gating signals for VSC. The3Ø-APF-PLL-based control with VSC provides abatement of harmonics, power support (real &reactive), load balancing, fast convergence, and improves the PQ of an islanded electrical supply system under various operational circumstances. The SPV system is modeled with an IC-MPPT algorithm using a boost converter at the DC-link of VSC to provide real/reactive power compensation necessary to regulate frequency and PCI voltage. The BES system supplies active/reactive power under the irregular condition of power generation sources and absorbs extra power during unbalanced load conditions. The 3Ø-APF-PLL-based SPV-wind-BES islanded electrical supply system has been designed in MATLAB/Experimental environment, and results are examined under various operational circumstances feeding unbalanced nonlinear load is discussed. The THD of PCI voltage (v_{gabc}) & current (i_{gabc}) are less than 5%, which satisfies the IEEE-519 standard also.

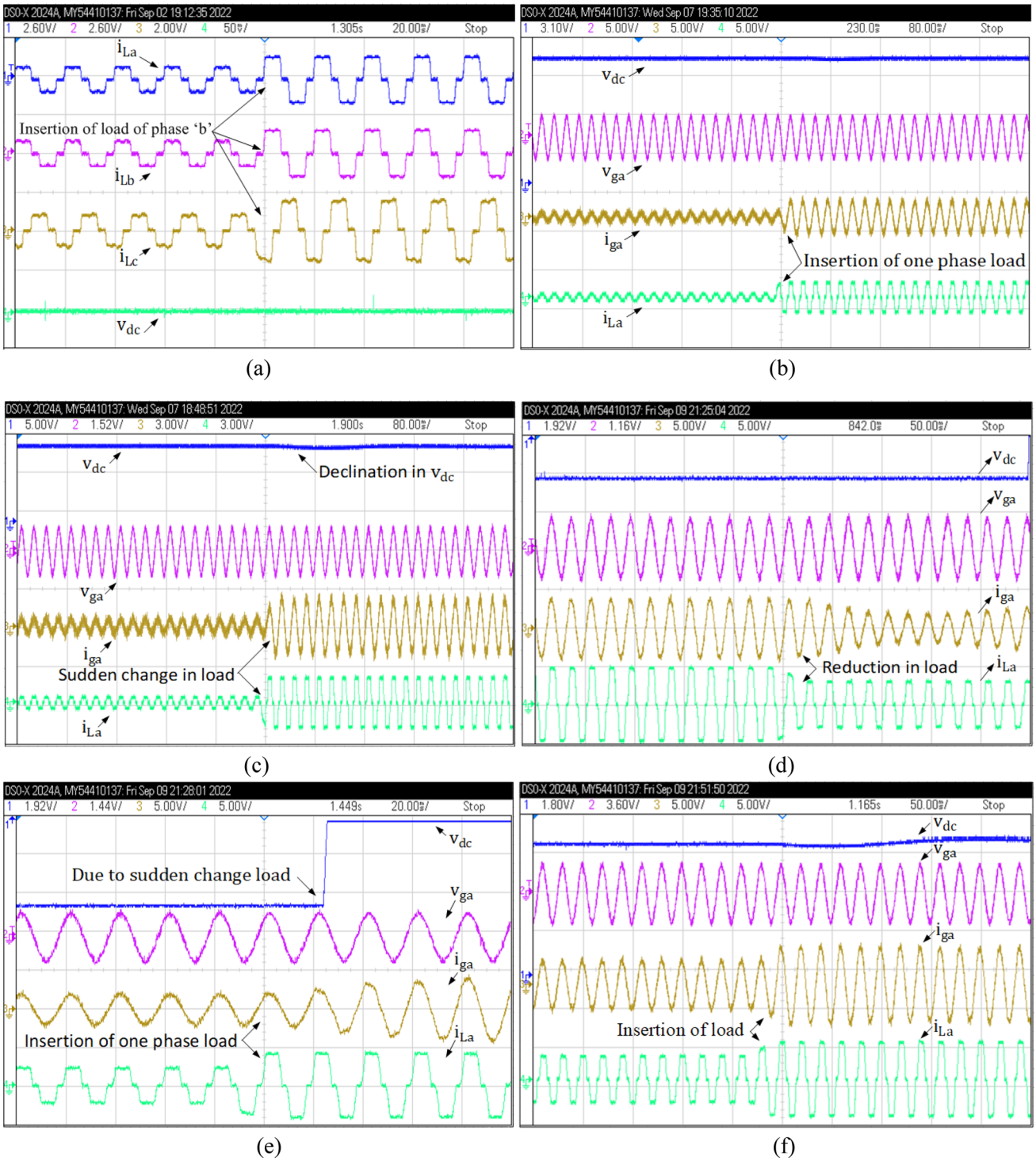


Fig. 11 **a–f** Dynamic behavior of islanded electrical supply system

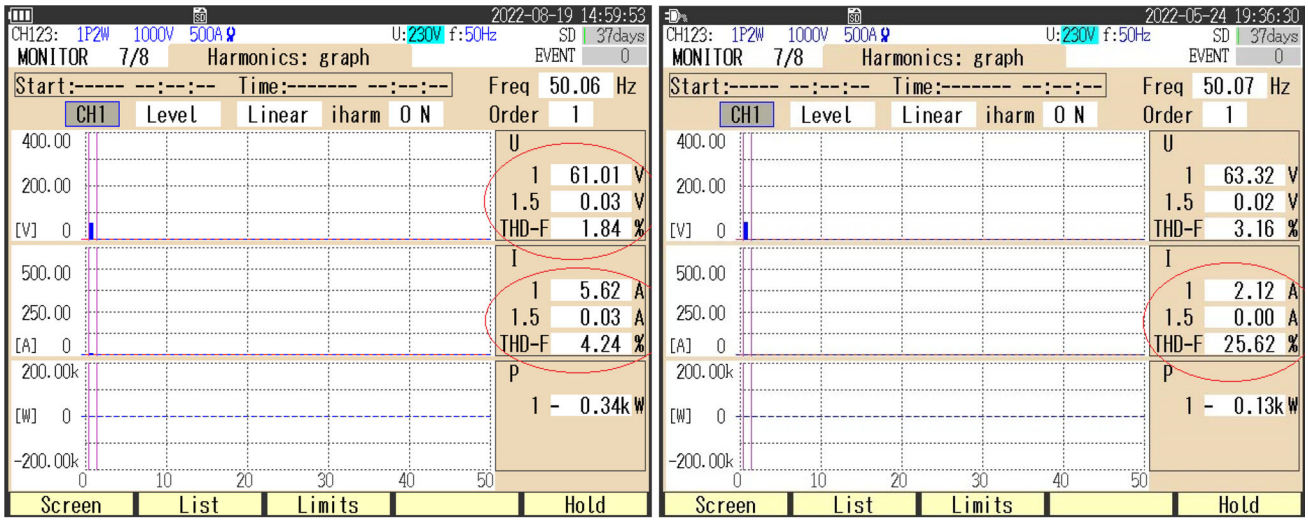


Fig. 12 **a** and **b** Harmonic spectrum

Author contributions "Sombir Kundu wrote the main manuscript text and prepared figures. All authors reviewed the manuscript.

Declarations

Conflict of interest Not applicable.

Appendix

Simulation and hardware parameters of wind & solar-based islanded electrical supply system.

S.No	Quantity	Parameters
1	SEIG and filtering inductors	3-Phase, 400 V, 3.7 kW, 4-Poles, 50 Hz and star connected $R_s = 1.405\ \Omega$, $L_s = 0.0078$ H, $R_r = 1.39$, $L_r = 0.0078$ H, Inertia = 0.138, Friction factor = 0, Pole Pairs = 2 and $L_f = 5$mH
2	PV system	$v_{pv} = 252.32$V, $i_{pv} = 11.69$A
3	Battery system	400 V, 7.5Ah
4	Nonlinear load	$R = 22\Omega$ and $L = 200\mu$H with 3-phase diode rectifier
5	Gain parameters	3Ø-APF-PLL: $k_p = 130.1$, $k_i = 7014.1$, $\omega_p = 157.1$rad/s,$\omega_s = 628.3$rad/s PI controller- in a terminal voltage control loop, $k_p = 0.4$ and $k_i = 0.05$, in frequency control loop, $k_p = 0.077$ and $k_i = 0.011$.
6	Hardware parameters	$v_{LL} = 110$V,50 Hz,$v_{dc} = 200$V, $k_p = 0.2$, $k_i = 0.01$, Nonlinear load: $R = 4\text{K}\Omega$ and $L = 100$mH with 3-phase diode rectifier

References

1. Sombir S, Singh M (2020) Voltage and frequency control of self excited induction generator integrated with PV system. In: IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society (pp. 4306-4311). IEEE. https://doi.org/10.1109/iecon43393.2020.9254608
2. Lopes LA, Almeida RG (2006) Wind-driven self-excited induction generator with voltage and frequency regulated by a reduced-rating voltage source inverter. IEEE Trans Energy Conv 21(2):297–304. https://doi.org/10.1109/tec.2006.874244
3. Singh B, Murthy SS, Chilipi RR, Madishetti S, Bhuvaneswari G (2014) Static synchronous compensator-variable frequency drive for voltage and frequency control of small-hydro driven self-excited induction generators system. IET Gener Transmission Distrib 8(9):1528–1538. https://doi.org/10.1049/iet-gtd.2013.0703
4. Kewat S, Singh B (2020) Grid synchronization of WEC-PV-BES based distributed generation system using robust control strategy. IEEE Trans Ind Appl 56(6):7088–7098. https://doi.org/10.1109/tia.2020.3021060
5. Puchalapalli S, Tiwari SK, Singh B, Goel PK (2020) A microgrid based on wind-driven DFIG, DG, and solar PV array for optimal fuel consumption. IEEE Trans Ind Appl 56(5):4689–4699. https://doi.org/10.1109/tia.2020.2999563
6. Benali A, Khiat M, Allaoui T, Denai M (2018) Power quality improvement and low voltage ride through capability in hybrid wind-PV farms grid-connected using dynamic voltage restorer. IEEE Access 6:68634–68648. https://doi.org/10.1109/access.2018.2878493
7. Kundu S, Singh M, Giri AK (2021) Design and control of a standalone wind–solar system with SEIG feeding linear/nonlinear loads. In: Proceedings of Symposium on Power Electronic and Renewable Energy Systems Control: PERESC 2020 (pp 317-327). Springer Singapore. https://doi.org/10.1007/978-981-16-1978-6_28
8. Pradhan S, Singh B, Panigrahi BK, Murshid S (2018) A composite sliding mode controller for wind power extraction in remotely located solar PV–wind hybrid system. IEEE Trans Ind Electron 66(7):5321–5331. https://doi.org/10.1109/tie.2018.2868009
9. Chauhan PJ, Chatterjee JK (2018) A novel speed adaptive stator current compensator for voltage and frequency control of standalone SEIG feeding three-phase four-wire system. IEEE Trans Sustain Energy 10(1):248–256. https://doi.org/10.1109/tie.2018.2868009
10. Arthishri K, Kumaresan N, Gounden NA (2018) Analysis and application of three-phase SEIG with power converters for supplying single-phase grid from wind energy. IEEE Syst J 13(2):1813–1822. https://doi.org/10.1109/JSYST.2018.2875761
11. Roodsari BN, Nowicki EP (2018) Analysis and experimental investigation of the improved distributed electronic load controller. IEEE Trans Energy Conv 33(3):905–914. https://doi.org/10.1109/tec.2018.2823334
12. Diab AAZ, Sultan HM, Mohamed IS, Kuznetsov ON, Do TD (2019) Application of different optimization algorithms for optimal sizing of PV/wind/diesel/battery storage stand-alone hybrid microgrid. IEEE Access 7:119223–119245. https://doi.org/10.1109/access.2019.2936656
13. Chauhan PJ, Reddy BD, Bhandari S, Panda SK (2018) Battery energy storage for seamless transitions of wind generator in standalone microgrid. IEEE Trans Ind Appl 55(1):69–77. https://doi.org/10.1109/tia.2018.2863662
14. Rezkallah M, Singh S, Chandra A, Singh B, Tremblay M, Saad M, Geng H (2019) Comprehensive controller implementation for Wind-PV-Diesel based standalone microgrid. IEEE Trans Ind Appl 55(5):5416–5428. https://doi.org/10.1109/tia.2019.2928254
15. Kundu S, Singh M, Giri AK (2022) Implementation of variable gain controller based improved phase locked loop approach to enhance power quality in autonomous microgrid. Int J Num Model Electron Netw Dev Fields. https://doi.org/10.1002/jnm.3082
16. Golestan S, Guerrero JM, Vasquez JC, Abusorrah AM, Al-Turki Y (2018) A study on three-phase FLLs. IEEE Trans Power Electron 34(1):213–224. https://doi.org/10.1109/tpel.2018.2826068
17. Kumar A, Kewat S, Singh B, Jain R (2020) CC-ROGI-FLL based control for grid-tied photovoltaic system at abnormal grid conditions. IET Gener Transmiss Distrib 14(17):3400–3411. https://doi.org/10.1049/iet-gtd.2019.0765
18. Mishra S, Hussain I, Pathak G, Singh B (2018) dPLL-based control of a hybrid wind–solar grid connected inverter in the distribution system. IET Power Electron 11(5):952–960. https://doi.org/10.1049/iet-pel.2017.0491
19. Golestan S, Guerrero JM, Vasquez JC (2019) Is using a complex control gain in three-phase FLLs reasonable? IEEE Trans Ind Electron 67(3):2480–2484. https://doi.org/10.1109/tie.2019.2903748
20. Giri AK, Arya SR, Maurya R, Babu BC (2019) VCO-less PLL control-based voltage-source converter for power quality improvement in distributed generation system. IET Electr Power Appl 13(8):1114–1124. https://doi.org/10.1049/iet-epa.2018.5827
21. Singh B (2016) Analysis and control of a standalone PV-BESS-DG based microgrid. In: 2016 7th India International Conference on Power Electronics (IICPE) (pp 1-6). IEEE. https://doi.org/10.1109/iicpe.2016.8079483
22. Chandran VP, Kewat S, Singh B (2020) Multi-objective control and operation of grid connected small hydro-solar PV-battery energy storage based distributed generation. IET Renew Power Gener. https://doi.org/10.1049/iet-rpg.2020.0212
23. Saxena H, Singh A, Rai JN (2020) Design and performance analysis of improved Adaline technique for synchronization and load compensation of grid-tied photovoltaic system. Int Trans Electr Energy Syst 30(6):e12388. https://doi.org/10.1002/2050-7038.12388
24. Ranjan A, Kewat S, Singh B (2020) Reweighted L 1 norm penalized LMS fourth algorithm of solar grid interfaced system for alleviating power quality problems. IEEE Trans Ind Appl 56(5):5352–5362. https://doi.org/10.1109/tia.2020.3004746
25. Singh S, Kewat S, Singh B, Panigrahi BK (2020) Enhanced momentum LMS-based control technique for grid-tied solar system. IET Power Electron 13(13):2767–2774. https://doi.org/10.1049/iet-pel.2019.1126
26. Sharma R, Kewat S, Singh B (2020) Robust MMSOGI-FLL control algorithm for power quality improvement of solar PV-SyRG pico hydro-BES based islanded microgrid with dynamic load. IET Power Electron 13(13):2874–2884. https://doi.org/10.1049/iet-pel.2019.1002
27. Kundu S, Singh M, Giri AK (2022) Control algorithm for coordinated operation of wind-solar microgrid standalone generation system. Energy Sour Part A Recov Utilization Environ Effects 44(4):10024–10044. https://doi.org/10.1080/15567036.2022.2143943

28. Ashutosh KG, Sabha RA, Dmitri V (2022) Distributed energy systems: design, modeling, and control, CRC press, (ed 1), p 308 https://doi.org/10.1201/9781003229124.
29. Kewat S, Singh B, Hussain I (2018) Power management in PV-battery-hydro based standalone microgrid. IET Renew Power Gener 12:391–398. https://doi.org/10.1049/iet-rpg.2017.0566
30. Singh B, Sharma R, Kewat S (2021) Robust control strategies for SyRG-PV and wind-based islanded microgrid. IEEE Trans Ind Electron. https://doi.org/10.1109/tie.2020.2978723
31. Golestan S, Guerrero JM, Vasquez JC, Abusorrah AM, Al-Turki Y (2020) All-pass-filter-based pll systems: linear modeling, analysis, and comparative evaluation. IEEE Trans Power Electron 35(4):3558–3572. https://doi.org/10.1109/TPEL.2019.2937936

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/373160133

Television as a Tool of Memory and Identity in Khaled Hosseini's Novels

Article *in* IAFOR Journal of Arts & Humanities · August 2023

DOI: 10.22492/ijah.10.1.04

CITATIONS
0

READS
11

2 authors:

Rajiv Ranjan Dwivedi
Delhi Technological University
10 PUBLICATIONS **2** CITATIONS

SEE PROFILE

Priya Meena
Delhi Technological University
1 PUBLICATION **0** CITATIONS

SEE PROFILE

All content following this page was uploaded by Priya Meena on 23 August 2023.

The user has requested enhancement of the downloaded file.

Television as a Tool of Memory and Identity in Khaled Hosseini's Novels

Priya Meena
Delhi Technological University, India

Rajiv Ranjan Dwivedi
Delhi Technological University, India

Abstract

Globalization epitomizes the convergence of diverse cultures, economies, and communities on a worldwide scale, engendering the exchange of ideas, technology, inventions, and import and export services across nations and societies. Technology, as a pivotal agent, assumes a vital role in building bridges among the global economies by overcoming information disparities and thereby empowering people. Specifically, television, a very influential medium, assumes the function of capturing, collating, and disseminating personal and cultural memories. Its significance is especially pronounced in the context of Khaled Hosseini's literary works, *The Kite Runner* and *A Thousand Splendid Suns*, where television's role as a tool for constructing memories and shaping the identities of the characters is scrutinized. The ramifications of the Taliban's television ban are examined in this paper, highlighting how this drastic measure has significantly impacted the nation's culture.

Keywords: American movies, culture, identity, memory, television

The credit for the advent of electronic television is unequivocally attributed to Philo Taylor Farnsworth, who successfully demonstrated its functionality in 1927 at a momentous event held in San Francisco, United States of America. The ensuing decades witnessed the proliferation of television's popularity across numerous countries, including the United Kingdom, Australia, Japan, the Soviet Union, China, and also India, which marked a significant milestone with its first TV channel, Doordarshan. It was only in the aftermath of the 1970s that countries like South Africa, Qatar, Oman, Burma, and Sri Lanka joined the ranks of television accessibility.

In the context of Afghanistan, the launch of television broadcasting occurred in 1978 with the establishment of Afghanistan National Television, a development catalyzed by the resolute advocacy of Dr. Hafiz Sahar, who recognized the indispensability of television as a tool for education and technological advancement. However, the path to progress, in terms of technology and utilization, was fraught with challenges for Afghanistan. Following its transition to a republic in the early 1970s, the country was beset by the intrusion of the Soviet Union in December 1979, a turbulent period that endured until 1989. The subsequent period of civil war from 1992 to 1996 further disrupted the nation's economy, culminating in the rise of the Taliban, an ideological group espousing Islamic Fundamentalism, which seized control.

Despite brief periods of resurgence and growth under NATO protection from 2001 onwards, Afghanistan's technological trajectory experienced a setback in 2021 as the Taliban once again assumed control, imperiling existing technology and depriving the citizens of their rightful access to knowledge and education. This unfortunate turn of events casts a somber pall over the lives of ordinary people, denying them the rights they so deserve.

In the present era, the ubiquity of television has entrenched itself as an indispensable aspect of modern society, profoundly influencing our outlooks, norms, and recollections. However, within the Taliban regime, the once taken-for-granted entitlement of television viewership faces scrutiny due to the imposition of their dogmatic Islamic ideology. Khaled Hosseini, a gifted diaspora writer, one renowned for his poignant portrayals of life's vicissitudes in Afghanistan, captures the intricate interplay of human relationships in a nation ravaged by conflict.

Curiously, despite the profundity of television's presence as a narrative device in Hosseini's literary works, scant scholarly attention has been devoted to exploring its multifaceted significance in his novels. Notably, television assumes a significant role in depicting emotional details in the lives of his characters, both within Afghanistan and in their adopted homelands, where they survive as refugees. Within this context, the pages of Hosseini's masterpieces, *The Kite Runner* and *A Thousand Splendid Suns*, offer a profound exploration of the diverse roles television plays in preserving memories, disseminating cultural values, and fostering connections between individuals separated by geographical expanses.

In this paper I endeavor to undertake a qualitative analysis, delving into the nuanced role of television in Hosseini's literary opus. Specifically, the focus will rest upon discerning how this medium functions as a potent instrument for preserving memory and shaping identities among the characters inhabiting Hosseini's narrative landscapes. By unfurling the myriad subtleties of television's thematic involvement in the tapestry of his stories, this inquiry aspires to provide insights on the profound intellectual and cultural implications intricately woven into the heart of Hosseini's timeless prose.

Television, Memory and Conflict

In his scholarly work, Rawan (2002) expounds on the widely acknowledged disapproval among orthodox followers of Islam concerning realistic depictions of living beings, perceiving photography and visual representation as an imitation of the divine creations. Such mimetic endeavors are believed to attract severe retribution from Allah on the Day of Judgment, even though the Quran itself does not overtly prohibit the portrayal of human forms (p. 162). Remarkably, the censure of graphic human depictions endures to this day, with contemporary religious adherents and clerics upholding the stance (p. 162).

Osman, in his book *Television and the Afghan Culture Wars* (2020), casts a light upon the deeply contentious nature of television within the Afghan context. This medium has historically been considered a crucible of violence due to its representation of social, cultural, and religious values. Paradoxically, television also has the potential to buttress democracy, national unity, and peace (p. 13). Furthermore, it has traditionally served as a conduit for Afghans to engage with foreign cultures, thereby exposing them to Western influences. Nevertheless, such exposure is diametrically opposed to the fundamental tenets of the Taliban's ideology, which staunchly imposes draconian restrictions on women's rights, encompassing prohibitions on public display of their bodies and faces and the curtailment of educational liberties. Under the Taliban regime, all forms of entertainment are rigorously prohibited, as they ostensibly contravene the conservative principles of Sharia Law (p. 13).

According to Rashid's account (1999), the tumultuous 1990s in Afghanistan were marked by a media ban, encompassing photography and television, which left the majority of the populace oblivious to the appearance of their newly instated leaders (p. 24).

Hatef and Cooke (2020) have propounded that a substantial segment of Afghanistan's populace lacks formal education. Consequently, television emerges as a more accessible and compelling medium for disseminating educational content vis-à-vis print media. The authors contend that television's wider reach holds the potential to furnish educational opportunities to a more extensive cross-section of Afghanistan's population.

In the aftermath of the 2001 triumph over the Taliban, Afghanistan underwent a transformative phase, metamorphosing into a nascent modern nation. An article from 2010 reveals that radio and television have become indispensable sources of entertainment for the Afghan people, with estimates suggesting that eight out of ten Afghans possess a radio and four out of ten possess a television (Auletta).

Amy Holdsworth's tome, *Television, Memory and Nostalgia*, delves into diverse forms of reflection that prove valuable in scrutinizing the interplay between television memory, the self, the family, domestic space, and the external world. Television, rather than being perceived as a mere object occupying a corner of the room, is seen as an integral facet of the domestic environment, engendering memories constructed from audiovisual impressions. It assumes a pivotal role in our comprehension of the past, and by attentively examining how it recollects and contextualizes its own history, we can discern prevailing attitudes towards television as a cultural phenomenon and towards our own historical identities. Nostalgia emerges as the dominant foundation through which television recollects and references itself.

Childhood Memories

In one of his interviews, Khaled Hosseini stated:

> I was watching a news story in the spring of 1999 on television, and this news story was about the Taliban. And it was talking about all the different impositions that the Taliban had placed on the Afghan people. And at some point along the line, it mentioned that they had banned the sport of kite flying, which kind of struck a personal chord for me, because as a boy I grew up in Kabul with all my cousins and friends flying kites. (RFE/RL)

Hosseini's literary voyage, kindled by the profound impact of televised news about Afghanistan, unveils television's powerful capacity to evoke long-lost childhood reminiscences. In his inaugural novel *The Kite Runner*, Hosseini artfully explores the characters' profound introspections as they retrace their past and establish emotional connections with their dear ones, catalyzed by the medium of television.

Set against the backdrop of a tumultuous Afghanistan marred by Russian invasions, the novel chronicles the poignant tale of Amir and his devoted friend and servant, Hassan. Amidst the harsh circumstances, Amir's father forges a path to America, taking Amir and leaving Hassan behind with aspirations for a brighter future. Years later, upon learning of Hassan's untimely demise, Amir embarks on a journey back to Afghanistan. Once there, he confronts his inner vulnerabilities and endeavors to rescue Hassan's son, all the while being engulfed by a wave of nostalgic recollections from his youth alongside his cherished companion, Hassan.

Permeated by poignant feelings, Amir's memories include his father's passion for soccer and his ritualistic pilgrimages to Tehran merely to witness "world cup games on television, since at the time Afghanistan didn't have TVs yet" (p. 19). This emotional reflection allows Amir to peer into the depths of his relationship with his father, unearthing a kaleidoscope of powerful emotions from that era of his life. The television, in its symbolic significance, serves as an emissary of memories intertwined with Hassan, epitomizing the profound bond that was their friendship.

The transformative era when Daoud Khan, Afghanistan's president from 1973 to 1978, heralded progress for the nation, notably with the advent of television, leaves an indelible mark on Amir's psyche. In contemplating the future acquisition of vividly colorful televisions, one each for himself and Hassan, Amir conveys his feeling of anticipation: "And not the black-and-white kind either. We'll probably be grown-ups by then, but I'll get us two. One for you and one for me." (p. 54). But fate's grim power thrusts the two kindred souls apart, ensnared in the perilous grip of their homeland's woes.

As destiny reunites Amir with Hassan's son, Sohrab, the melancholic sight of the young boy transfixed before the television, with "his green eyes, stone-faced, rocking back and forth," strikes an evocative chord within Amir's soul (p. 286). The television acts as a conduit, ferrying Amir back to the memories of his cherished camaraderie with Hassan and, concurrently, stirs within him the unfulfilled promises that linger as haunting echoes of his past.

Hosseini's eloquent portrayal of the interplay between television and memory weaves a tapestry of emotions, bridging the chasm between past and present, thereby underscoring the timeless human struggle for connection and redemption.

A Dreamland of Prosperity

Hosseini's literary works abound with allusions to American cinema, which constitutes an integral facet of the cultural milieu. In a particular recollection of Amir's, he divulges the denouement of *The Magnificent Seven* to a boy who inquired if he had viewed the film. This incident, transpiring at the video store, imparted a crucial lesson to Amir: in the United States, the disclosure of endings is deemed inappropriate, a divergence from Afghanistan's customs, where inquiries about denouements, particularly concerning the attainment of happiness and fulfillment of aspirations by characters, are commonplace. This preoccupation with happiness mirrors the yearnings harbored by the Afghan populace. Nonetheless, Amir, in the course of reminiscing, comes to the realization that life rarely adheres to cinematic conventions, particularly when confronted with the desolate and traumatized landscape of Afghanistan, as is illustrated in one of the chapters: "We passed the rusted skeleton of an overturned car, a TV set with no screen half-buried in rubble" (p. 232). These movies function as a conduit between South Asian nations such as Afghanistan, where the quest for happiness is fervent, and the Western world, where the United States embodies a land of affluence and the realization of dreams and desires. Despite America – his newfound home – having buried his childhood memories, the sight of the Pacific Ocean nearly moves Amir to tears, evoking recollections of his conversations with Hassan about visiting the beach and relishing the sensation of water beneath their feet, having hitherto only encountered "oceans on the movie screens" (p. 126).

It is noteworthy that Afghanistan, ensconced by mountains, is landlocked, rendering the experience of an awe-inspiring ocean view nearly unattainable. An article on thestar.com recounts the reminiscences of a man who lived under the Taliban's rule during his formative years, a period in which all forms of entertainment were strictly proscribed. Haris and his neighbor pooled their resources to acquire a pirated copy of the film *Titanic*, an expenditure amounting to "ten times as much as a night's rental" (Latifi). After a month of anticipation, they finally had Haris's home to themselves, where they watched the film. Haris shares an astonishing reaction to witnessing the majestic boat on-screen, stating, "I couldn't believe what I was seeing, they actually created a boat. I remember everyone was saying the actual boat was the size of Khair Khāna [a neighborhood in north west Kabul]" (Latifi). The allusion to the sea in *A Thousand Splendid Suns* emerges when Hosseini mentions that the movie Titanic encapsulates themes of "the sea. The luxury. The ship" (p. 296). Hosseini's portrayal not only accentuates the unattainable nature of beholding the sea, emblematic of vastness and freedom, luxury representing a life of abundance, and the ship as a means of escape, but also underscores the myriad challenges routinely faced by the impoverished Afghans.

The Making of an Identity

In his masterpiece *A Thousand Splendid Suns*, Khaled Hosseini unfolds the poignant narrative of Laila, a modern and well-educated Afghan woman, whose life takes a tragic trajectory following the loss of her parents, the separation from her beloved Tariq, and her ensnarement in an abusive marriage with the elderly and brutal Rasheed, who is already wed to Mariam.

Initially, a sense of animosity prevails between Mariam and Laila, yet the shared subjugation they endure as women under the oppressive grasp of Rasheed forces an unlikely convergence, fostering a profound sisterhood between them. Throughout the novel, we bear witness to the relentless rule of the Taliban, their unremitting cruelty impelling the characters to seek liberation from their harrowing circumstances.

Remarkably, the medium of television assumes a prominent role in shaping the identities of the characters. Laila's affinity for television and cinema serves as a lens through which we observe the evolution of her character. Notably, the allure of cinema conjures reminiscences of Tariq, as well as recollections of her father, underscoring the intricate interplay of personal connections and familial heritage in the author's construal of the cultural significance of cinema.

Osman aptly expounds on the influence of television in Afghanistan, elucidating its dual function as both a mirror and an amplifier of the nation's essence. By enabling the Afghan populace to engage in self-observation and articulate their own imagery and representations, television assumes an integral role in shaping the nation's collective consciousness.

For Laila, her experiences and relationships with her loved ones profoundly color the memories and emotional associations she harbors with cinema. Beyond a mere source of entertainment, cinema assumes a deeply personal and culturally significant dimension for her, molding her sense of self and contributing to the construction of her individual identity.

Upon receiving the disconcerting news of the Taliban's cinema closures, Laila's mind drifts back to the days spent in her father's cinema park, where she fondly reminisces about the moments shared with Tariq, indulging in the captivating narratives of Hindi movies. These poignant tales of star-crossed lovers, torn asunder by the capricious whims of fate, one adrift in distant lands, the other ensnared in the shackles of forced matrimony, evoking tears and serenades amidst fields of marigolds, all resonate deeply with the somber realities of her own distressing life.

Despite the stark contrast between the illusory realm of cinematic dreams and the relentless hardships that mark her existence, these cherished movie memories serve as a consoling balm, an avenue of escapism for Laila, offering a glimmer of optimism and the promise of a brighter tomorrow amid the gloom. Ultimately, destiny intertwines their paths, and in the denouement, Laila finds solace in a long-awaited reunion with Tariq, reaffirming the enduring power of love and hope in the face of adversity.

Amidst the aura of a promising and idyllic reunion shared by Laila and Tariq, they soon confront a stark disparity between the realities they now face and the idealized portrayals found in the romantic films of yesteryears. Departing from each other has wrought profound changes, evident in the melancholic countenances they take on while witnessing President George W. Bush declaring war on the Taliban. In this poignant moment, Tariq's voice falters, and Laila senses the weight of his emotions threatening to manifest as tears.

Their agonizing struggle to reclaim the euphoria of cherished memories, despite their relentless troubles, underscores not only the transformations in the external world but also the profound metamorphoses within themselves. The televised declaration of war serves as an unwelcome trigger, casting Laila's thoughts back to the tragic fate her parents endured - a cruel demise at the hands of a rocket during the Afghan Civil War.

The declaration, too, serves as a poignant trigger, rekindling somber memories of Laila's brothers, Ahmad and Noor, who fell victims to the ravages of war. When Tariq ventures the notion that the conflict may not necessarily be adverse, it stirs a maelstrom of emotions within Laila, underscoring the intricate interplay of their perspectives.

Holdsworth astutely notes the potent influence of news and media events in constructing lasting imprints in our television memories, encapsulated by the dictum, "You'll remember where you were and who you were with the first time you saw them" (2011, p. 244). The profound impact of the war announcement in the novel leaves an indelible mark on both Laila and Tariq, forging memories that will forever be etched in their consciousness.

A parallel scenario emerges in *The Kite Runner*, where an adult Amir, while watching "the communist propaganda" and witnessing the unfolding of "the second war" on his television screen, is forever bound to those memories that linger like an enduring specter. The recollection of encountering the Taliban for the first time on the small screen, amid the allure of Afghanistan's landscape upon his return, demonstrates the enduring impression cast by televised imagery.

Hosseini's perspicacious delineation of the "Titanic fever" phenomenon in Afghanistan, triggered by the screening of the eponymous movie, serves as a captivating exploration of the profound sociocultural ramifications that media consumption had under the Taliban's rule. The surreptitious dissemination of pirated copies of the film, discreetly concealed even in the recesses of one's underclothing amidst the restrictive curfew, conjures a poignant tableau where spectators surreptitiously shed tears for the fateful protagonists, Jack, Rose, and the ill-fated passengers aboard the ill-fated vessel.

This shared encounter bestows upon the collective psyche an indelible memory, transcending the confines imposed by an oppressive regime. By partaking in this clandestine cinematic communion, the people of Afghanistan forge a poignant interconnection that defies the suppressive measures of authority and fosters a profound cultural resilience amidst adversity.

In Laila's psyche, the tale of Jack and Rose begets touching reminiscences of her own bygone romance with Tariq. She discerns striking resemblances between the tragic fate of Jack and that presumed to befall Tariq. Both figures, she firmly believes, have succumbed to an untimely demise. This realization profoundly contributes to her despondent state and a pervasive sense of despair, for she apprehends that neither Jack nor Tariq shall ever come to her rescue. As she succinctly proclaims, "Every soul seeks salvation in a Jack-like figure amidst catastrophe, but such a savior remains a mere chimera—Jack shall not return, for Jack is no more." (p. 297)

The scenario wherein Mariam and Laila's daughter, Aziza, engage in a theatrical representation inspired by the movie epitomizes their longing for escapism. Aziza, in a pronounced act of identification, assumes the persona of Jack, while Mariam takes on the role of Rose. When Mariam, in jest, remarks that Aziza shall perish young, while she herself will live to a ripe old age, Aziza resolutely counters that her demise shall be heroic akin to Jack's. And when Laila's son Zalmai queries which part he shall assume, Aziza audaciously assigns him the character of "the iceberg." (p. 296) This fervent aspiration to embody heroism, rather than subscribing to the conventional, subordinate female archetype of being rescued, especially amid the women's collective yearning for a Jack-like savior amidst Afghanistan's oppressive circumstances, unequivocally showcases Aziza's courage and nonconformist disposition. The scene, thus, accentuates her capacity to find solace and amusement amid the yoke of oppressive conditions, thereby enriching her individuality. Furthermore, this lighthearted enactment facilitates the forging of cherished memories amongst the trio.

For Mariam, Laila, and the children, *Titanic* assumes the role of an escapist medium, much like it does for other Afghans. Limited access to resources in Afghanistan often restricts their

possibilities, compelling them to immerse themselves in fantasies of a serene existence beyond their borders. However, television furnishes a window into the outside world, offering glimpses of life in other nations and imbuing them with a sanguine outlook towards their own existence. The power to immerse themselves in an alternate realm through television can prove to be a potent source of solace and inspiration for Afghans, providing a brief respite from their daily hardships and instilling in them the impetus to strive for a brighter future. Jaya Yadav aptly underscores the role of television as an escape from reality in Hosseini's works, noting that despite Afghanistan's stringent prohibitions against "illicit" materials, films, television sets, and VCRs were clandestinely smuggled, providing many Afghans a portal to escape their harsh actuality (p. 867).

Moreover, the juxtaposition of the American cartoon movie "Pinocchio" within the novel *A Thousand Splendid Suns* affords a perceptive lens through which to explore the intricacies of Mariam's character, establishing a profound metaphor for her profound sense of isolation and displacement. Much like Pinocchio, the anthropomorphic figure crafted from wood and imbued with life, Mariam emerges as a child of unwedded lineage, reared in destitution within a diminutive village, despite being the progeny of an affluent father. This milieu engenders a persistent sense of alienation and a recurrent disdain from the villagers, akin to Pinocchio's struggle to find his place within the human realm. Consequently, Mariam's longing for acceptance and affection becomes distinctly molded by this animated cinematic narrative.

Amidst the allegorical fabric lies Mariam's father's broken promise to attend the same film, an act that betrays his own deceitful nature, paralleling the gradual elongation of Pinocchio's nose with each falsehood he utters. This unfulfilled pledge serves as a potent symbol for her father's negligence and the estrangement that pervades their relationship, mirroring the detachment evident in Pinocchio's own familial dynamics. Both characters seem to perpetuate a profound feeling of disconnection, exacerbating their shared sense of isolation.

While Mariam's father belatedly acknowledges the harm inflicted upon his daughter through rejection and disavowal, the revelation comes too late to mend their fractured bond. An earnest effort to atone materializes in the form of a movie dispatched to Mariam, yet tragically, it never finds its way to her. This celluloid creation, an inextricable facet of Mariam's life, is unveiled as Walt Disney's *Pinocchio* during a pivotal scene where Laila endeavors to forge a connection with Mariam's past. Laila's persistent scrutiny of the film, alternating between playback and rewind, inadvertently illustrates the profound significance it holds in Mariam's existence. Alas, despite Laila's sincere endeavors, she attempts to fully grasp the movie's profound import in Mariam's psyche.

Ultimately, akin to the metamorphosis of Pinocchio, Mariam experiences a transformative realization of her intrinsic worth, ultimately embracing her authentic self in the denouement.

In the denouement of the novel, a particular American film, *Superman*, causes Zalmai's bemusement because of the curious notion that, in Afghanistan, a boy would be precluded from bearing the appellation "Clark" (p. 402). Shahzad (2011) delves into an intriguing examination of a respondent's post-television viewing experience, reminiscent of Zalmai's inquisitive nature, wherein the participant transcends a mere passive observer or impotent bystander of the encoded contextual tapestry. Rather, they actively exercise agency, making discerning choices and probing inquiries that pertain to the embedded messages within the representations (p. 383).

Conclusion

The utilization of television in Khaled Hosseini's literary works serves as a discerning portrayal of its role as a medium for evoking memories and bridging the cultural chasm between Afghanistan and the West. In *The Kite Runner*, the recollections of Amir's formative years with his companion, Hassan, become intertwined with the advent of television in Afghanistan. Additionally, references to American movies within the novel facilitate a connection between the two disparate cultures.

Similarly, in *A Thousand Splendid Suns*, television plays a momentous role in shaping Laila's character, enabling her to forge connections with memories of her father, Babi, and Tariq. It serves as a reflective and amplifying lens for Afghanistan, allowing for introspection and communication of diverse representations. These novels exemplify how television has the potential to evoke intimate and cultural ties, even amidst the backdrop of calamitous circumstances.

Subsequent to the Taliban's invasion in 2021, the survival of media, particularly television, has been imperiled. Despite international efforts, the domestic television broadcasts have depicted how female journalists were initially mandated to conceal their faces, and eventually, they were altogether prohibited from appearing on screen. This jeopardizes the future and masks the heritage of Afghanistan, as opportunities for exposure to other cultures are once again constrained, impeding people's ability to connect with the outside world. The concerns extend beyond just the fate of women but also encompass the freedom of expression for all individuals through television.

Nevertheless, there have been instances of resolute defiance. Notably, a Kabul University Professor displayed a powerful act of protest by tearing up his diplomas on live television, signifying his disapproval of the deteriorating system under the new leadership. This demonstration not only underscores the influence of this medium but also emboldens people to resist the encroaching despondency. As he vociferously articulated, "From today, I don't need these diplomas anymore because this country is no place for an education. If my sister and my mother can't study, then I don't accept this education." (Bose, 2022)

The impact of television on Afghan society is complex, defying uniformity and predictability. Its effects are intricately woven with a multitude of factors, encompassing the political, economic, and cultural dimensions, necessitating future comprehensive analyses and scholarly investigations.

References

Auletta, K. (2010, June 28). The networker Afghanistan's first media mogul. *The New Yorker*. https://www.newyorker.com/magazine/2010/07/05/the-networker-2

Bose, S. K. (2022, December 28). "If mother, sister can't...": Afghan professor tears up diplomas on TV. *NDTV*. https://www.ndtv.com/world-news/taliban-afghanistan-ban-on-women-education-if-mother-sister-cant-afghan-professor-tears-up-diplomas-on-tv-3643516

Hatef, A., & Cooke, T. R. (2020). Winning hearts and Mminds: A critical analysis of independent media development in Afghanistan. *Journal of International and Intercultural Communication, 13*(2), 114–129. https://doi.org/10.1080/17513057.2020.1740764

Holdsworth, A. (2011). *Television, Memory and Nostalgia*. Palgrave Macmillan.

Hosseini, K. (2007). *A Thousand Splendid Suns*. Bloomsbury.

Hosseini, K. (2003). *The Kite Runner*. Bloomsbury.

Latifi, A.M. (2022, December 17). How James Cameron's Titanic became an under-the-surface sensation in Taliban-ruled Afghanistan. *The Toronto* Star. https://www.thestar.com/news /world/2022/12/16/how-james-camerons-titanic-became-an-under-the-surface-sensation-in-taliban-ruled-afghanistan.html

Osman, W. (2020). Television and the Afghan culture wars: Brought to you by foreigners, warlords, and activists. Urbana: University of Illinois.

Radio Free Europe/Radio Liberty. (2012, June 21). 'Kite Runner' Author on his childhood, his writing, and the plight of afghan refugees. *RFE/RL*. https://www.rferl.org/a/interview-kite-runner-afghan-emigre-writer-khaled-hosseini/24621078.html

Rashid, A. (1999). The Taliban: Exporting extremism. *Foreign Affairs, 78*(6), 22–35. https://doi.org/10.2307/20049530

Rawan, S. M. (2002). Modern mass media and traditional communication in Afghanistan. *Political Communication, 19*(2),155–170. http://dx.doi.org/10.1080/10584600252907425

Shahzad, F. (2011). Collective memories: A complex construction. *Memory Studies, 5*(4), 378º391. http://dx.doi.org/10.1177/1750698011419121

Yadav, J. (2019). Reading Lolita in Tehran and a thousand splendid suns: Through history, politics, gender. *International Journal of Advanced Research, 7*(10), 863–869. http://dx. doi.org/10.21474/IJAR01/9901

Corresponding Author: Priya Meena
Email: priya1996meena@gmail.com

View publication stats

Int J Syst Assur Eng Manag
https://doi.org/10.1007/s13198-023-02077-x

ORIGINAL ARTICLE

The effectiveness of research and development tax incentives in India: a quasi-experimental approach

Aakanksha Kaushik[1]

Received: 5 September 2022 / Revised: 24 July 2023 / Accepted: 25 July 2023
© The Author(s) under exclusive licence to The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2023

Abstract Globally, governments provide significant tax incentives to business enterprises for research and development (R&D) expenditure to foster innovation. Several fiscal incentives are provided to firms registered under India's department of scientific and industrial research (DSIR). India provided a super deduction of 200% on R&D expenses in 2011, reduced to 150% in 2016 and 100% in 2020. The study uses firm-level data to test the effectiveness of incentives offered in India and analyze the impact of the rationalization of super deduction. For the study, the difference in difference analysis is done to evaluate the changes in the outcome variables: total, current, and capital R&D expenditures. The treatment group is the firms registered under DSIR. The results show a significant impact of the mix of tax incentives provided on both current and capital expenditures. The super deduction significantly impacted only current expenses and not capital expenditures. The administrative costs under current expenditures are easier to relabel than capital expenditures. The study shows that the overly generous regime incentivized firms to relabel their non-R&D costs as R&D expenses for profiteering. The study supports the move to reduce the super deduction.

Keywords R&D · Tax incentives · Quasi-experiment · Difference-in-differences

JEL Classification H26 · H32 · O3

✉ Aakanksha Kaushik
aakanksha.kaushik@dtu.ac.in

[1] University School of Management and Entrepreneurship, Delhi Technological University, Delhi, India

1 Introduction

Business enterprises' expenditures on Research and Development (R&D) play a critical role in promoting innovation. Joseph Schumpeter described innovation as "Creative Destruction" and identified it as a key to economic growth (Schumpeter 1962). Additionally, Endogenous growth theories have emphasized the significance of innovation and technological advancement for growth. (Lucas 1988; Romer 1990). Recent studies have also shown that innovation promotes firm-level competitiveness (Bacinello et al. 2020; Qiu et al. 2020; Suat and San 2019). A vast literature has shown that innovation increases productivity (Griliches 1958, 1980; Mansfield 1965; Griliches et al. 1984; Mansfield 1984; Mun et al. 1991; Nadiri and Kim 1996; Frantzen 1998; Aghion and Howitt 1998; Klette & Kortum 2002).

Firms' lack of investment in R&D is caused by financial limitations and variations in social and private returns. As a result, governments have been motivated to step in and offer incentives to encourage more R&D investment. (Nelson 1959; Arrow 1962; Aghion and Howitt 1992; Jaffe 1996; Hausmann et al. 2003; Rodrik 2004a, b.)

Many countries use R&D tax incentives as policy tools. These incentives can lower the cost of R&D, and are known as input-based incentives. Examples include Tax Deferrals, Allowances, Credits, Concessional Import Tariffs, GST, and Accelerated Depreciation. Other incentives can increase profits earned from R&D, such as Rate Relief, Tax Holidays, and the Patent Box Regime. Several R&D tax incentives in India are available to firms registered under the Department of Scientific & Industrial Research (DSIR). If eligible, firms not registered under DSIR can still claim an income tax deduction on R&D expenses and benefits from the patent box regime and tax holidays. For a summary of tax incentives provided by India and recent changes, see Table 1.

Published online: 21 August 2023

Table 1 Recent changes in R&D tax incentives

S. no.	Incentives	Recent changes
1	Customs duty exemptions	Union budget 2022–23 announced immediate withdrawal of basic customs duty (BCD) exemptions on goods used for Research and development in agro-chemical sector unit
		End-date of 31.03.2023 is prescribed in Union Budget 2022–23 for goods used in the pharmaceutical and biotechnology sector
		All the entries in the notification 50/96-Customs dated 23.07.1996, unless varied or rescinded, will have validity up to 31.03.2023
		Continues
2	Income tax deduction	The rate of weighted tax deduction has been lowered from 200 to 150% from 1st April 2017, and from Annual Year 2021–22, the tax deduction is 100% of the expenditure incurred
3	Accelerated depreciation	Continues
4	Concessional GST rate	Continues
5	Patent box regime	It was introduced in 2016
6	Tax holiday	A one-year extension to the start-ups incorporated till March 31, 2023, was provided under Union Budget 2022–23. The incentive was introduced in 2016

After the pandemic, evaluating the cost of tax incentives is important. The Indian government has made some changes to their policies, which are listed in Table 1. Companies registered under DSIR can receive benefits such as Weighted Tax deduction, Concessional Customs duty, Concessional GST Rate, and Accelerated Depreciation. However, they must fulfill certain requirements such as setting up R&D centers with specific areas and employees, as well as keeping audited accounts of R&D expenditures. One major incentive is the income tax-weighted deduction of 200% on R&D expenditures, also known as the Super deduction. Firms can receive an income tax deduction on R&D expenditures under Sect. 35 of the IT Act, 1961. India has a broader definition of qualifying expenses for exemption compared to other countries, with all current and capital expenses being eligible except for land or building costs. The deduction has changed over time, with it being revised to 150% in 2000–01, 200% in 2010–11, 150% in 2016–17, and finally reduced to 100% from the annual year 2021–22. The study aims to test how effective the incentives provided under DSIR are and evaluate the impact of the rationalization of the Super deduction.

The Super Deduction tax incentive for R&D expenses has the potential downside of encouraging companies to mislabel their non-R&D administrative costs as R&D to take advantage of the 200% tax deduction. This issue has been identified in other countries as well, such as in studies by Mansfield and Switzer (1985), GAO (1989, 2009), Bloom et al. (2019), and Chen et al. (2021). This paper aims to investigate whether Indian firms also engage in this practice of relabelling expenses. While no previous studies have examined this issue in India, this paper fills that gap by providing new evidence on relabelling practices. The quasi-experimental approach of difference-in-difference analysis is used to address potential data endogeneity issues.

The rest of this paper is organized as follows: Sect. 2 provides an overview of the research background and hypothesis development. Section 3 describes the data and methodology. Section 4 presents the findings. Finally, Sect. 5 concludes with some reflections on the implications of our results.

2 Research background and hypothesis development

There are two main types of studies that evaluate the effectiveness of tax incentives: those that measure the impact on individual companies (micro level) and those that measure the impact on a larger scale (macro level). The macro-level studies use overall research and development spending as the dependent variable and examine the effects of tax incentives at the national and state levels. Researchers such as Bloom et al. (2002), Athukorala and Kohpaiboon (2010), McKenzie and Sershun (2010), Moretti and Wilson (2017), and Brown et al. (2017) have conducted studies in this area.

At the micro level, studies have been conducted to evaluate the effects of R&D tax incentives on firms. These studies measure the impact of such incentives on a company's production of new products or processes patents (output additionality) and on its R&D expenditures (input additionality) (Cappelen et al. 2008; Czarnitzki et al. 2011; Colombo et al. 2011, Ivus et al. 2021). The input additionality studies use micro econometric analysis techniques to analyze the impact of R&D tax or generosity of tax on the R&D expenditures of the firm (Berger 1993; Hall and van Reenen 2000; Bloom et al. 2002; Corchuelo and Martinez-Ros 2009; Czarnitzki et al. 2011; Duguet 2010). Some studies have used surveys for R&D data (Oxeraa 2006; Czarnitzki et al. 2011; Cappelen et al. 2012; Lokshin and Mohnen 2012; Lokshin and

Table 2 Number of firms recognised under the DSIR over the years. Source: Own compilation from various annual reports of the department of scientific & industrial research (DSIR)

	2011–2012	2012–2013	2014–2015	2015–2016	2016–2017	2017–2018	2018–2019	2019–2020	2021–2022
In-house R&D units recognition	1618	1767	1762	1800	1880	1997	2052	2238	2481
New recognition	151	149	105	161	183	163	234	186	141

Mohnen 2013; Mulkay and Mairesse 2013; Guceri 2018) and some studies used annual reports and administered corporate return data for analysis (Yang et al. 2012; Rao 2016; Dechezleprêtre et al. 2016; Guceri and Liu 2019; Agrawal et al. 2020; Chen et al. 2021; Ivus et al. 2021). For this particular study, R&D data from audited annual reports of firms in the country were obtained from the CMIE prowess database.

Several studies have focused on the issue of relabelling expenses (Eisner et al. 1984; Mansfield and Switzer 1985; GAO 1989, 2009; Guceri 2018; Bloom et al. 2019; Chen et al. 2021). The US GAO conducted studies in 1989 and 2009, highlighting concerns over the definition of qualifying expenses and their monitoring. This issue is particularly problematic for developing countries. A recent study by Chen et al. 2021 found that relabelling accounts for 24.2% of reported R&D. It's important to keep in mind that obtaining DSIR certification involves maintaining a separate record of R&D expenses that Statutory Auditors have verified. The firms must also provide information about their projects to renew their certificates periodically. Administrative expenses can only be reclassified as current expenditures to a limited extent. A meta-analysis of microeconomic studies has revealed that R&D tax incentives have a greater impact on current expenditures than capital expenditures (Ladinska et al. 2015). This study aims to determine the effectiveness of R&D tax incentives in stimulating capital and current expenditures while also checking for any signs of reclassification.

One of the main issues with analyzing data is the presence of endogeneity problems, which occur when companies choose to participate in incentive programs on their own. This is also the case in India, where firms self-select and register under DSIR to receive benefits. Due to this concern, several studies use a direct approach to analysis, including Regression Discontinuity, Matching Analysis, or Difference-in-Difference Analysis (Hægeland and Møen 2007; Corchuelo and Martínez-Ros 2009; Yohei 2011; Yang et al. 2012; Bozio et al. 2014; Agrawal et al. 2014; Rao 2016; Dechezleprêtre et al. 2016; Bronzini and Piselli 2016; Guceri 2018; Wang 2018; Ivus et al. 2021). In this study, we use the difference in difference analysis methodology based on a recent survey of incentives in India by Ivus et al. (2021). However, there are two main differences from their study. Firstly, our data covers a longer period, from 2011 to 2021, and focuses on the impact of reducing the super deduction from 200 to 150% on R&D expenditures and patents. Secondly, we test the effects on current and capital expenditures separately, as no previous study has examined the presence of relabelling in India. Our main contribution is to provide evidence of relabelling in the country.

3 Data and methodology

This study primarily utilizes data extracted from two sources: the CMIE prowess database and the DSIR directory of in-house R&D units. The CMIE prowess database contains information on individual companies, gathered from audited annual reports, the Ministry of Company Affairs, and company filings with stock exchanges. This database includes data on the firms' capital, current and total R&D expenditures, and other financial information. The yearly data of firms registered with the DSIR is obtained from the department's annual DSIR directory. Table 2 displays the number of firms registered under DSIR and new recognitions from 2011 to 2021.

In this paper, we examine the R&D spending of companies in two categories: financial and non-financial. The non-financial category comprises firms from various major industries such as manufacturing, mining, electricity, construction, real estate & irrigation, and services. The study collects extensive data from a diverse range of companies operating within these categories. Most of the analyzed companies belong to the Drugs & Pharmaceutical and Computer Software industry groups.

The treatment dummy is the DSIR status of the ith firm in the jth year and is denoted by D_{it}. The variable takes value one if the firm i is registered with the DSIR in year j; otherwise, the value is zero. This divides the sample into two groups. The treated group is the firms registered under DSIR, and the control group is the group of firms not registered under DSIR. The variable C_t is the year when the super deduction incentive was available and takes value one for 2011–2016. So, the interaction term $D_{it}C_t$; allows for the difference in the impact of the change in tax incentive on R&D expenditure across two groups.

Table 3 Outcome and control variable

Outcome variables	
Log of total R&D intensity	Log of (total R&D expenditures/sales)
Log of capital R&D intensity	Log of (capital R&D/sales)
Log of current R&D intensity	Log of (Current R&D/sales)
Control variables	
Age	Number of years since operation
Exporter	A dummy variable = 1 if firm is an exporter
Importer	A dummy variable = 1 if firm imports capital goods
Leverage	Total borrowings divided by total assets of the firm
Firm size	Log of gross value of fixed assets

In this study, we use the variable D_{it} to represent the DSIR status of the ith firm in the jth year. This variable takes on a value of one if the firm is registered with DSIR in that year and zero otherwise. Using this variable, we divide our sample into two groups: the treated group, consisting of firms registered under DSIR, and the control group, consisting of firms not registered under DSIR.

Additionally, we use the variable C_t to represent the year when the super deduction incentive was available. This variable takes on a value of one, from 2011 to 2016. By creating an interaction term between $D_{it}C_t$, we can analyze the difference in the impact of the change in tax incentive on R&D expenditure across the two groups.

The difference-in-difference specification of the model is as follows:

$$LogRD_{it} = \alpha + \alpha_t + \beta D_{it} + \gamma D_{it}C_t + \delta X_{it}^{'} + \epsilon_{it} \quad (1)$$

where RD_{it} is one of the three measures of ith firm's R&D intensity in jth year, which is calculated by dividing the R&D expenditures by the firm's sales. The three outcome variables are total R&D intensity, capital R&D intensity, and current R&D intensity. α_t denotes the time-fixed component in the model. The percentage difference between DSIR registered and non-DSIR firms is $(e^{\beta-1})100$. The coefficient γ captures the differential change before 2016 for the treatment group compared to the control group. The percentage treatment effect of the change is $(e^{\gamma-1})100$. The X_{it} is a vector with a size of K × 1. Various control variables of the vector include Technology Imports, Firm Size, Leverage, importer dummy, exporter dummy, and age (which are listed in Table 3). The vector also includes a constant and ϵ_{it}, which represents a stochastic constant term. Table 4 shows the summary statistics of these variables.

4 Findings

Out of the 1244 firms included in the sample, 378 were registered under the DSIR consistently between 2011 and 2021, while 519 were never registered. The status of DSIR registration for all other firms fluctuated over time.[1] The model estimation results are presented in two sections. The first section analyzes the effects on companies that have been registered under DSIR from 2011 to 2021. In Sect. 4.2, the results are studied after taking into account the variable that changes over time.

4.1 Constant DSIR status

The information presented in this sample is based on data from 378 companies that were registered under DSIR from 2011 to 2021. The model was applied to 6276 observations, with results displayed in Table 5. Column 1 shows the total R&D expenditure intensity, column 2 displays the current R&D expenditure intensity, and column 3 presents the capital R&D expenditure. The variables of interest are the treatment variable D_{it} and the interaction variable $D_{it}C_t$.

Table 4 Summary statistics of variables

Variable	Mean	Std. dev	Min	Max
Total R&D expenditure	51.30	196.59	0.00	2924.50
Current R&D expenditure	42.07	163.87	0.00	2310.13
Capital R&D expenditure	10.70	73.75	0.00	1748.20
Total R&D expenditure intensity	0.02	0.07	0.00	3.64
Current R&D expenditure intensity	0.01	0.05	0.00	2.23
Capital R&D expenditure intensity	0.00	0.02	0.00	1.25
Technology imports	0.01	0.03	0.00	0.42
Firm size	6.08	2.00	4.61	13.56
Leverage	1.07	9.13	0.00	484.21
importer	1.00	0.05	0.00	1.00
exporter	0.48	0.50	0.00	1.00
Age	36.08	22.67	0.00	158.00

[1] For example, A company named '20 Microns Nano Minerals Limited' was registered under DSIR till 2018 but not afterwards. There were many such companies and therefore model with the time-varying treatment variable (DSIR registration status) is estimated.

Table 5 Constant DSIR status results

Outcome variable in logs	(1) Total R&D expenditure intensity	(2) Current R&D expenditure intensity	(3) Capital R&D expenditure intensity
Treatment group	1.417***	1.396***	0.460***
	(0.111)	(0.114)	(0.152)
Interaction effect	0.286***	0.364***	0.137
	(0.033)	(0.027)	(0.145)
Age	−0.0127	−0.012***	−0.012***
	(0.001)	(0.001)	(0.002)
Size	−0.20472***	−0.23186***	−0.203***
	(0.009)	(0.009)	(0.017)
Leverage	0.00429*	0.00442*	−0.002
	(0.000)	(0.000)	(0.000)
Importer	0.713	0.739*	−0.617
	(0.634)	(0.513)	(2.286)
Exporter	−0.208*	−0.255**	0.160
	(0.118)	(0.104)	(0.380)
Constant	−5.083***	−5.101***	−4.827***
	(0.103)	(0.091)	(0.323)
Year fixed effects?	Yes	Yes	Yes

***Significant at 1% level, **5% level, *10% level. Robust Standard errors are in brackets

The data in column (1) indicates that the beta coefficient for variable D_{it} is positive ($\beta = 1.417$) and the gamma coefficient for variable $D_{it}C_t$ is also positive ($\gamma = 0.286$). Both coefficients are highly statistically significant, which means that the average Total R&D expenditure intensity in DSIR-registered firms is 4.1 times higher than in non-registered firms. Additionally, the estimated interaction effect is 1.33, meaning that the treatment effect of super deduction is 33%.

The second column of results shows that the coefficient β for the variable D_{it} is 1.396, and the coefficient γ for the variable $D_{it}C_t$ is 0.364. Both coefficients are positive and highly statistically significant, similar to the results for Total R&D expenditures. This means that the average Current R&D expenditure intensity in DSIR-registered firms' treatment groups is 4.03 times greater than the control group of non-registered firms. Additionally, the estimate of the interaction effect is 1.44, indicating that the treatment effect of super deduction is 44%.

Based on the data presented in column (3), it can be observed that the coefficient β is 0.460 for the variable D_{it}, which is a positive and highly significant value. Meanwhile, the coefficient γ is 0.137 for the variable $D_{it}C_t$, but this value is statistically insignificant. This means that the average capital R&D expenditure intensity in treatment groups of DSIR registered firms is 1.58 times greater than the control group of non-registered firms, as determined by the $e^{0.460}$ value. However, the impact of the super deduction treatment on capital R&D expenditures is not significant.

4.2 Time-varying DSIR registration status

Table 6 displays the outcome when the DSIR status, which is the treatment variable, varies over time. The findings in Table 6 are comparable to those in Table 5. The estimates are elevated when the DSIR status is time-varying, but the conclusions are the same. Companies that are registered under DSIR have invested more in research and development (R&D), but the 200% super deduction has only significantly affected current expenses and not capital expenses.

The first column of results indicates that the coefficient β for the variable D_{it} is positive, with a value of 1.833. The coefficient γ for the variable $D_{it}C_t$, is also positive, with a value of 0.252. These coefficients are both highly statistically significant. This suggests that the average Total R&D expenditure intensity is 6.25 times greater for treatment groups of DSIR-registered firms compared to the control group of non-registered firms. The estimate for the interaction effect is 1.29, implying that the treatment effect of a 200% super deduction is 29%.

The second column of results shows that the coefficients for the variables D_{it} and $D_{it}C_t$ are both positive and highly statistically significant. Specifically, β is 1.826 and γ is 0.325. This indicates that the average Current R&D expenditure intensity in treatment groups of DSIR-registered firms is 6.21 times greater than that of the control group of non-registered firms. The interaction effect estimate is 1.38, meaning that the 200% super deduction treatment effect is 38%.

Table 6 Time-varying DSIR status results

Outcome variable in logs	(1) Total R&D expenditure intensity	(2) Current R&D expenditure intensity	(3) Capital R&D expenditure intensity
Treatment group	1.833***	1.826***	0.658***
	(0.109)	(0.114)	(0.188)
Interaction effect	0.252***	0.325***	0.042
	(0.026)	(0.021)	(0.365)
Age	−0.011***	−0.010***	−0.012***
	(0.001)	(0.001)	(0.002)
Size	−0.191***	−0.216***	−0.187***
	(0.008)	(0.008)	(0.019)
Leverage	0.006***	0.007***	−0.004
	(0.001)	(0.001)	(0.000)
Importer	0.512	0.522	−0.601
	(0.645)	(0.543)	(1.692)
Exporter	−0.270***	−0.306***	0.142
	(0.078)	(0.076)	(0.632)
Constant	−5.676***	−5.705***	−5.287***
	(0.076)	(0.065)	(0.333)
Year fixed effects?	Yes	Yes	Yes

***Significant at 1% level, **5% level, *10% level. Robust Standard errors are in brackets

According to the information in column (3), the β coefficient for variable D_{it} is 0.658, which is both positive and highly significant. However, the coefficient for variable $D_{it}C_t$ is $\gamma = 0.042$ and statistically insignificant. This means that the average capital R&D expenditure intensity for DSIR-registered firms in treatment groups is 1.93 times greater than that of non-registered control groups. However, the impact of the 200% super deduction treatment on capital R&D expenditures is insignificant.

In this study, the expenses of companies were examined in two categories: financial and non-financial. The model was evaluated for both categories, and the results revealed that non-financial firms had significant impacts while financial firms did not. It is important to note that financial firms made up less than 10% of the sample size. The study also highlighted that non-financial firms play a greater role in the overall research and development expenditure of the economy.

5 Conclusion

This research paper aims to address the topic of relabelling, which has limited literature available. Several governments have changed the R&D tax incentives policy in India from 2011 to 2021. One of the significant changes was the reduction of income tax deductions from 200% in 2022 to 150% in 2016 and eventually to 100% in 2021. This presents an opportunity to examine the effectiveness of the generous deduction and other incentives given to firms registered under DSIR. These firms can avail of several fiscal incentives. The study considers the firms registered under DSIR as the treatment group and uses the difference-in-difference approach to test their effectiveness. The research assesses the changes in R&D expenditure during the period when the Super deduction was available.

The study found that there was a clear difference in the average R&D spending between the companies in the treated group and those in the control group. However, while the super deduction offered was very generous, it only significantly impacted current R&D spending and not capital R&D spending. This lack of impact on capital R&D spending suggests that some companies may have tried to relabel their spending. The increase in Total R&D spending was driven solely by current spending. The super deduction caused an increase of 44% in current spending but had no significant effect on capital R&D spending. This suggests that some companies may have increased their current spending in order to benefit from the tax deduction without increasing their capital spending at a commensurate rate.

The DSIR offers a combination of incentives that effectively increase current and capital R&D expenditures. The fiscal incentives provided by DSIR have resulted in a six-fold increase in current expenditures and almost a two-fold increase in capital expenditures. However, it is important to note that while incentives to encourage R&D expenditures are beneficial, overly generous schemes can lead to abuse

of the system. The Super deduction, for example, incentivizes firms to falsely label administrative expenses as R&D expenditures to take advantage of tax deductions without any actual increase in innovation.

Funding No funds, grants, or other support was received.

Declarations

Conflict of interest I declare that there are no financial and non-financial conflicts of interest.

Human or animal rights The Research does not involve Human Participants and/or Animals.

Informed consent Not applicable.

References

Aghion P, Howitt P (1992) A model of growth through creative destruction. Econometrica 60:323–351

Aghion P, Howitt P (1998) A Schumpeterian perspective on growth and competition. In: Coricelli F, Matteo MD, Hahn F (eds) New theories in growth and development. Palgrave Macmillan

Agrawal A, Rosell C, Simcoe T (2020) Tax credits and small firm R&D spending. Am Econ J Econ Pol 12(2):1–21

Agrawal A, Rosell C, Simcoe TS (2014) Do tax credits affect R&D expenditures by small firms? Evidence from Canada, NBER Working Paper 20615

Arrow KJ (1962) Economic welfare and the allocation of resources to invention. The rate and direction of inventive activity. Princeton University Press

Athukorala PC, Kohpaiboon A (2010) Globalization of R&D by US-based multinational enterprises. Res Policy 39(10):1335–1347

Bacinello E, Tontini G, Alberton A (2020) Influence of maturity on corporate social responsibility and sustainable innovation in business performance. Corp Soc Responsib Environ Manag 27(2):749–759

Berger PG (1993) Explicit and implicit tax effects of the R & D tax credit. J Accounting Res 31(2):131–171

Blandinieres F, Steinbrenner D, Weiß B (2020) Which design works? A meta-regression analysis of the impacts of R&D tax incentives. A meta-regression analysis of the impacts of R&D tax incentives, 20-010

Bloom N, Griffith R, Van Reenen J (2002) Do R&D tax credits work? Evidence from a panel of countries 1979–1997. J Public Econ 85(1):1–31. https://doi.org/10.1016/S0047-2727(01)00086-X

Bloom N, Van Reenen J, Williams H (2019) A toolkit of policies to promote innovation. J Econ Perspect 33(3):163–184

Bozio A, Irac D, Py L (2014) Impact of the research tax credit on R&D and innovation: evidence from the 2008 French reform, Banque de France Working Paper 532

Bronzini R, Piselli P (2016) The impact of R&D subsidies on firm innovation. Res Policy 45(2):442–457

Brown JR, Martinsson G, Petersen BC (2017) What promotes R&D? Comparative evidence from around the world. Res Policy 46(2):447–462. https://doi.org/10.1016/j.respol.2016.11.010

Cappelen Å, Raknerud A, Rybalka M (2012) The effects of R&D tax credits on patenting and innovations. Res Policy 41(2):334–345. https://doi.org/10.1016/j.respol.2011.10.001

Chen Z, Liu Z, Suárez Serrato JC, Xu DY (2021) Notching R&D investment with corporate income tax cuts in China. Am Econ Rev 111(7):2065–2100. https://doi.org/10.1257/aer.20191758

Colombo MG, Luca G, Samuele M (2011) R&D subsidies and the performance of high-tech startups. Econ Letters 112(1):97–99

Corchuelo MB, Martínez-Ros E (2009) The effects of fiscal incentives for R&D in Spain, Universidad Carlos III de Madrid Working Paper 09-23

Czarnitzki D, Hanel P, Rosa JM (2011) Evaluating the impact of R&D tax credits on innovation: a micro econometric study on Canadian firms. Res Policy 40(2):217–229. https://doi.org/10.1016/j.respol.2010.09.017

Dechezleprêtre A, Einiö E, Martin R, Nguyen KT, Van Reenen J (2016) Do tax incentives for research increase firm innovation? An RD design for R&D (No. w22405). National Bureau of Economic Research

Duguet E (2010) The effect of the R&D tax credit on the private funding of R&D: An econometric evaluation on French firm level data. Available at SSRN 1592988

Eisner R, Albert SH, Sullivan MA (1984) The new incremental tax credit for R&D: incentive or disincentive? Natl Tax J 37(2):171–183

Frantzen D (1998) R & D, international technical diffusion and total factor productivity. Kyklos 51:489–508

Gao (1989) Tax policy and administration: the research tax credit has stimulated some additional research spending. U.S. Government Accountability Office, Washington

Gao (2009) Tax policy: the research tax credit's design and administration can be improved. U.S. Government Accountability Office, Washington

Griliches Z (1958) Research costs and social returns: hybrid corn and related innovations. J Polit Econ 76:141–154

Griliches Z (1980) Returns to research and development expenditures in the private sector. New developments in productivity measurement. University of Chicago Press, pp 419–462

Griliches Z, Mairesse J (1984) Productivity and R&D at the firm level. Am Econ Rev 84:66–83

Guceri I (2018) Will the real R&D employees please stand up? Effects of tax breaks on firm-level outcomes. Int Tax Public Financ 25(1):1–63. https://doi.org/10.1007/s10797-017-9438-3

Guceri I, Liu L (2019) Effectiveness of fiscal incentives for R&D: Quasi-experimental evidence. Am Econ J Econ Policy 11(1):266–291

Hægeland T, Møen J (2007) Input additionality in the Norwegian R&D tax credit scheme, Reports 2007/47 Statistics Norway

Hall B, Van Reenen J (2000) How effective are fiscal incentives for R&D? A review of the evidence.Res Policy 29(4–5):449–469

Hausmann R, Rodrik D (2003) Economic development as self-discovery. J Dev Econ 72(2):603–633

Ivus O, Jose M, Sharma R (2021) R&D tax credit and innovation: Evidence from private firms in India. Res Policy 50(1):104128. https://doi.org/10.1016/j.respol.2020.104128

Jaffe A (1996) Economic analysis of research spillovers: implications for the advanced technology program. Econ Anal 1:14

Klette T, Kortum S (2002) Innovating firms and aggregate innovation, NBER Working Paper 8819. National Bureau of Economic Research

Ladinska E, Non M, Straathof B (2015) More R&D with tax incentives? A meta-analysis (No. 309). CPB Netherlands Bureau for Economic Policy Analysis

Boris L, Mohnen P (2012) How effective are level-based R&D tax credits? Evidence from the Netherlands. Appl Econ 44(12):1527–1538

Lokshin B, Mohnen P (2013) Do R&D tax incentives lead to higher wages for R&D workers? Evid Neth Res Policy 42(3):823–830. https://doi.org/10.1016/j.respol.2012.12.004

Lucas R (1988) The mechanics of economic growth. J Monet Econ 22:3–42
Mansfield E (1965) Rates of return from industrial research and development. Am Econ Rev 55(1/2):310–322
Mansfield E (1984) R&D and innovation: some empirical findings. In: Griliches Z (ed) R&D, patents, and productivity. University Press for National Bureau of Economic Research, pp 127–154
Mansfield E, Switzer L (1985) The effects of R&D tax credits and allowances in Canada. Res Policy 14(2):97–107. https://doi.org/10.1016/0048-7333(85)90017-4
McKenzie KJ, Sershun N (2010) Taxation and R&D: an investigation of the push and the pull effects. Can Public Policy 36(3):307–324. https://doi.org/10.3138/cpp.36.3.307
Moretti E, Wilson DJ (2017) The effect of state taxes on the geographical location of top earners: evidence from star scientists. Am Econ Rev 107(7):1858–1903. https://doi.org/10.1257/aer.20150508
Mulkay B, Mairesse J (2013) The R&D tax credit in France: assessment and ex ante evaluation of the 2008 reform. Oxf Econ Pap 65(3):746–766
Mun HW (1991) Korea's total factor productivity, Seoul: Korean Productivity Center. https://doi.org/10.3386/w5506
Nadiri MI, Kim S (1996) International R&D spillovers, trade and productivity in major OECD countries
Nelson RR (1959) The simple economics of basic scientific research. J Polit Econ 67(3):297–306
Oxera (2006) Feasibility study for potential econometric assessment of the impact of R&D tax credits on R&D expenditure, HM Revenue and Customs Research Report 19, Oxford. http://www.innovation-policy.org.uk/compendium/reference/Default.aspx?referenceid=124
Qiu L, Jie X, Wang Y, Zhao M (2020) Green product innovation, green dynamic capability, and competitive advantage: Evidence from Chinese manufacturing enterprises. Corp Soc Responsib Environ Manag 27(1):146–165
Rao N (2016) Do tax credits stimulate R&D spending? The effect of the R&D tax credit in its first decade. J Public Econ 140:1–12. https://doi.org/10.1016/j.jpubeco.2016.05.003
Rodrik D (2004a) Industrial Policy for the Twenty-First Century, CEPR Discussion Paper 4767. Centre for Economic Policy Research
Rodrik D (2004b) Industrial policy for the twenty-first century. Available at SSRN 666808. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=666808
Romer P (1990) Endogenous technological change. J Polit Econ 98:71–102
Schumpeter J (1962) The theory of economic development. Springer
Suat LA, San OT (2019) Corporate environmental management: eco-efficiency and economics benefits among manufacturers certified with EMS14001 in Malaysia. Int J Recent Technol 7(6):873–886
Wang J (2018) Innovation and government intervention: a comparison of singapore and hong kong. Res Policy 47(2):399–412
Yang C-H, Huang C-H, Hou T-T (2012) Tax incentives and R&D activity: firm-level evidence from Taiwan. Res Policy 41(9):1578–1588
Yohei KOBA (2011) Effect of R&D tax credits for small and medium-sized enterprises in Japan: evidence from firm-level data, RIETI Discussion Paper

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

Third Order Quadrature Oscillator with Arbitrary Phase Shift Generation and Amplitude Modulation Using MDXCCCTA

Garima Shukla, Sajal K. Paul

***Abstract*— This work presents the realization of an electronically tunable third-order quadrature oscillator (TOQO) employing a current mode block, modified dual-X current controlled current conveyor transconductance amplifier (MDXCCCTA) and three grounded passive capacitors. The oscillator simultaneously provides two quadrature voltage outputs, two high-impedance quadrature current outputs, and one high-impedance current output with controllable amplitude. Further, the proposed circuit works as an arbitrary phase shift generator from 0 to $\pm 90^0$. In addition, the realization generates amplitude modulation (AM) or amplitude shift keying (ASK) signals with controllable amplitude without affecting the condition of oscillation (CO) and frequency of oscillation (FO). The CO and FO are electronically tunable and can be controlled independently. Moreover, an automatic control method is proposed to stabilize the CO and FO against the variation in relative amplitudes of quadrature signals when TOQO works as an arbitrary phase shift generator. It exhibits low phase error (0.22^0), low % THD (0.78-0.88)), low phase noise (-180 dBc to -240 dBc), low sensitivity (<0.45%) and low power consumption (1.08 mW). Moreover, the circuit is vastly robust against tolerances in passive components and process, voltage and, temperature (PVT) corners as the deviations from nominal values are small (<0.8%). Furthermore, the TOQO occupies comparatively a small layout chip area ($7810\mu m^2$). It is a compact module providing quadrature, arbitrary phase shifted, and AM signals from the same realization. Therefore, the proposed design can be a handy module for instrumentation, measurement, and communication systems. Experimental results are presented to verify simulation and theoretical analyses.**

***Index Terms*— AM/ASK, MDXCCCTA, Oscillator, Phase shift, Quadrature, Resistor less, Third Order.**

I. Introduction

Sinusoidal oscillators/ quadrature oscillators find wide applications in instrumentation, measurement, analog signal processing, and communication system. A quadrature oscillator (QO) generates two sinusoidal signals having 90^0 phase shifts. Measurement of an unknown capacitance using the principle of an oscillator is a well-investigated scheme in an instrumentation system as presented in [1, 2]. A QO realized in [3] has been used to measure the soil's impedance to characterize the soil's different properties, like humidity and salinity. It is well acknowledged that QOs generating multiple sinusoids with 90^0 phase shifts are important blocks for quadrature mixers and amplitude modulators (AM) in transmitter and receiver systems. There is a requirement for AM signals in various detection systems. An AM-based ultrasonic dynamic range detection method for distance measurement is illustrated in [4], where the transmitter employs a sinusoidal generator and analog multiplier as a modulator in the system. This detection system is also essential in very accurate quartz sensor methods such as [5, 6], where an oscillator and mixer are required in the measurement system of quartz crystal microbalance gas sensors.

Considering the diverse applications of QO, a wide variety of QOs has emerged [7-38] in the past few decades. An electronically tunable QO with enhanced tuneability is presented in [7]. A few arbitrary phase shift oscillator configurations are presented in [8].

During the past decades, the current-mode (CM) approach has been receiving considerable attention in analog integrated circuit design over the voltage mode counterparts due to their potential advantages, such as inherently wide bandwidth, higher slew rate, greater linearity, wider dynamic range, simple circuitry, and low power consumption [9]. The QO reported in [7] is a second-order circuit using a CM building block.

Moreover, the third-order quadrature oscillators (TOQOs) are immensely preferred over the second-order oscillators due to their higher accuracy, higher quality factor, lower harmonic distortion, and lower phase noise; as a consequence, the realization of several TOQOs with various CM building blocks have been reported [10-38]. The TOQO realizations in [10-30, 32, 33, 37, 38] generate quadrature signals only, and some of the TOQOs designed in [31, 33, 34, 35, 36] have the feature of analog modulation (AM/ASK) within them. Furthermore, various applications require quadrature current and quadrature voltage signals simultaneously. The TOQOs designed in [12, 15, 16, 17, 20-21, 22, 24, 25, 29, 31, 32, 33-38] simultaneously generate quadrature current and voltage signals. Moreover, oscillators with a variable FO without disturbing the CO are a desirable requirement in oscillator designs; however, independent controlling of FO and CO through electronic tuning are available only in [11, 15, 18, 19, 20-21, 26-28, 29-30, 33-37].

The TOQOs in [10-45] have been realized with two or more analog building blocks (ABBs), and [37, 38] are realized using only one ABB.

The exhaustive study of [7, 8, 10-38] as illustrated in Table V reveals the following shortcomings; (i) none of the realizations work as a QO, arbitrary phase shift generator, and

This paragraph of the first footnote will contain the date on which you submitted your paper for review, which is populated by IEEE.
(Corresponding author: Sajal K. Paul).
Garima Shukla and Sajal K. Paul are with Department of Electronics Engineering, Indian Institute of Technology (ISM), Dhanbad, 826004, India. (e-mail: garima29shukla@gmail.com; sajalkpaul@rediffmail.com).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

an AM generator from the same circuit topology, (ii) most of the oscillator circuits use an excessive number of active/passive components, (iii) no realization possessing full electronic tuneability and having small layout chip area provides quadrature currents (CM), quadrature voltages (VM), and AM signals simultaneously (iv) many of the circuit designs do not offer the independent controlling of FO and CO through electronic tuning (v) many of the realizations possess high % THD, high phase error, and high power consumption.

In view of attempting to fill some of the research gaps, this research presents the realization of a novel third-order quadrature oscillator (TOQO) employing a current mode building block, MDXCCCTA, and three grounded passive capacitors. Further, the same circuit works as an arbitrary phase and AM generator. The proposed TOQO has the following main advantageous features: (i) entirely electronically tunable circuit, (ii) provides two VM and two CM quadrature signals, simultaneously, (iii) independent controlling of CO and FO through separate electronically tunable parameters (bias currents), (iv) high impedance current outputs, hence suitable for cascading, (v) exhibits high frequency of oscillation (vi) works as an AM/ASK signal generator without affecting the CO and FO, (vii) works as an arbitrary phase shift generator/oscillator (APSO) providing arbitrary phase in the range, 0 to $\pm 90^0$, (viii) CO and FO are simultaneously controlled against variation in relative amplitudes by a new proposed automatic stabilization method implemented in proposed TOQO, (ix) variable frequency through electronic tuning, (x) wide electronically and non-electronically tunable range of frequency of oscillation, (xi) occupies comparatively small layout chip area, (xii) exhibits low phase error, low THD, low phase noise and low power consumption, (xiii) vastly robust against tolerances in passive components as verified by Monte Carlo simulation (xiv) small deviation against various process, voltage and, temperature corners as verified by PVT corner analysis, (xv) effect of parasitics of MDXCCCTA in circuit performance is negligible as confirmed by post layout simulation, and (xvi) low intermodulation distortion during AM generation.

Hence, the proposed topology consisting of QO, modulator (AM/ASK), and arbitrary phase shift generator may be useful as a compact module in measurement, instrumentation, and communication systems.

This paper is organized into seven sections. Section I is this section; herein, the introduction is given. Section II proposes the TOQO using MDXCCCTA and further its illustration as an arbitrary phase generator, and AM generator is given. Section III illustrates the automatic CO and FO control method in the proposed TOQO. The non-ideal analysis is illustrated in section IV. Section V presents simulation results with a discussion, followed by experimental results in Section VI. A comparison with the available literature is given in section VII, followed by the conclusion in section VIII.

II. Proposed Oscillator

This section discusses the MDXCCCTA, followed by the realization of a third-order quadrature oscillator (TOQO) using MDXCCCTA. The extension of TOQO as an arbitrary phase shift generator and amplitude modulation is also given.

A. MDXCCCTA

The second-generation current conveyor (CCII) is a versatile current mode building block. It has primarily three terminals, X, Y, and Z. The voltage at the X terminal is equal to the Y terminal, whereas the current at the X terminal is conveyed to the Z terminal with unity gain. To add flexibility to CCII, dual-X current conveyor (DXCCII) [39], followed by DXCCCII [40, 41], was introduced. Another version of DXCCII was reported as the Dual-X current conveyor transconductance amplifier (DXCCTA) [42-44]. The DXCCTA combines DXCCII and a transconductance amplifier (TA) [42-44]. In contrast, to achieve electronic control through the X± terminal's intrinsic resistances in DXCCCII, electronic tunability in DXCCTA is introduced by controlling TA's transconductance in DXCCTA.

Electronic tuneability through controlling the X± terminal's intrinsic resistance is one of the most desirable features in CM building blocks, such as in DXCCCII. However, no DXCCTA block has the electronic tuneability feature through controlling intrinsic resistances at X± terminals. This paper proposes a current controlled DXCCTA as a new variant of CCII, combining the features of DXCCCII and DXCCTA with one more cascaded TA. It may be named a modified Dual-X current controlled current conveyor transconductance amplifier (MDXCCCTA).

The circuit symbol and internal CMOS structure of the MDXCCCTA are shown in Fig. 1 and Fig. 2, respectively. As depicted in Fig. 2, the proposed MDXCCCTA is similar to conventional DXCCTA [42-44] except with the internal modification for the current controllability of intrinsic resistances at X± terminals. Fig. 2 shows the availability of an electronic tuneability feature for controlling intrinsic resistances, R_{X+} and R_{X-}, present at X+ and X- terminals through bias currents, I_{b1} and I_{b2}, respectively, and the availability of an additional tuning through bias current (I_{b4}) of second TA. The resistances at Y, Z, O_1, and O_2 terminals are high. The port relationships of MDXCCCTA in matrix form are given as follows;

$$\begin{pmatrix} I_Y \\ V_{X+} \\ V_{X-} \\ I_{Z+} \\ I_{Z-} \\ I_{O1+} \\ I_{O2+} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & R_{X+} & 0 & 0 & 0 \\ -1 & R_{X-} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & g_{m1} & 0 \\ 0 & 0 & 0 & g_{m2} & -g_{m2} \end{pmatrix} \begin{pmatrix} V_Y \\ I_{X+} \\ I_{X-} \\ V_{Z+} \\ V_{Z-} \end{pmatrix} \quad (1)$$

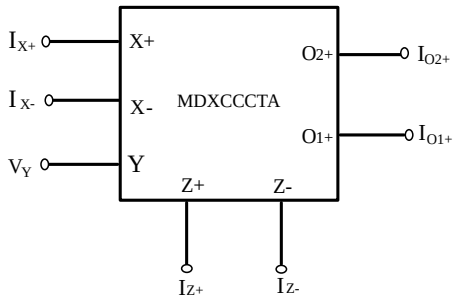


Fig. 1. Symbolic representation of MDXCCCTA.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

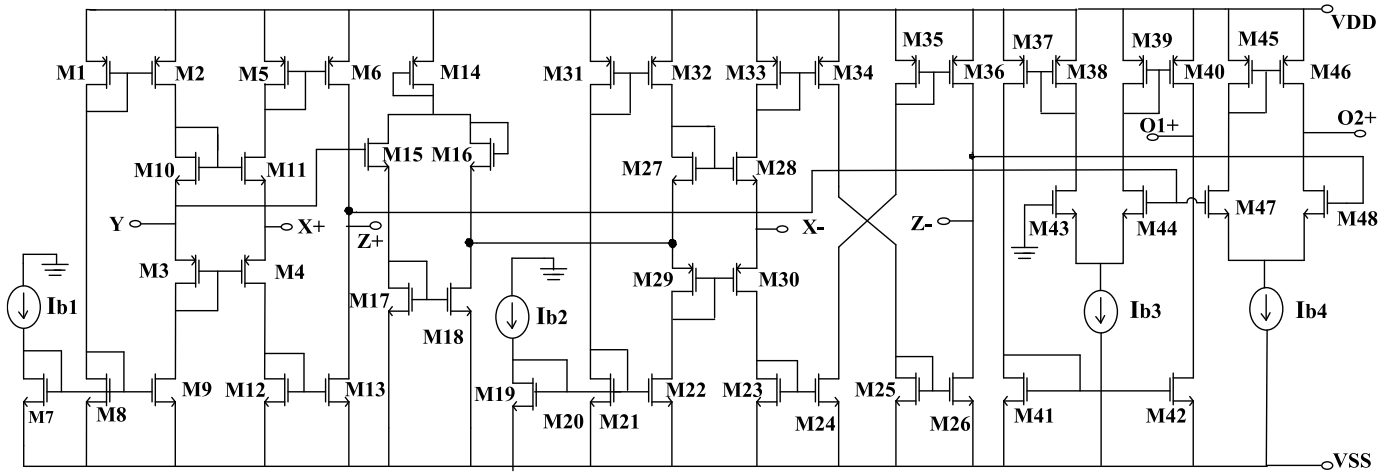


Fig. 2. Internal CMOS structure of MDXCCCTA.

Where in (1),

$$R_{X+} = \frac{1}{\sqrt{8k_1 I_{b1}}}, \quad R_{X-} = \frac{1}{\sqrt{8k_2 I_{b2}}} \tag{2}$$

Where, $$k_1 = \mu_n C_{OX}\left(\frac{W}{L}\right)_{10,\,11} = \mu_p C_{OX}\left(\frac{W}{L}\right)_{3,\,4} \tag{3}$$

$$k_2 = \mu_n C_{OX}\left(\frac{W}{L}\right)_{27,\,28} = \mu_p C_{OX}\left(\frac{W}{L}\right)_{29,\,30} \tag{4}$$

It may be observed from (2) that R_{X+} and R_{X-} can be tuned through the bias currents, I_{b1} and I_{b2}, respectively. The g_{m1} and g_{m2} are the trans-conductance gains of the first and second transconductance amplifiers (TAs) of MDXCCCTA and are expressed as;

$$g_{m1} = \sqrt{k_3 I_{b3}} \tag{5}$$

Where, I_{b3} is the bias current and $k_3 = \mu C_{OX}\left(\frac{W}{L}\right)_{43,\,44}$. (5a)

$$g_{m2} = \sqrt{k_4 I_{b4}} \tag{6}$$

Where, I_{b4} is the bias current and $k_4 = \mu C_{OX}\left(\frac{W}{L}\right)_{47,\,48}$.

B. Proposed Third-Order Quadrature Oscillator Circuit

1). Quadrature Signals Generator

The realization of the proposed TOQO is based on the principle of circuit connection of one current mode inverting lossy integrator and two current mode non-inverting lossless integrators, as shown in Fig. 3.

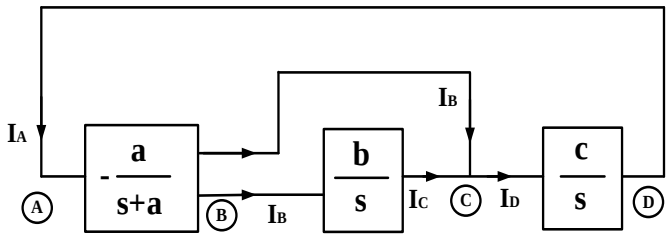


Fig. 3 Functional block diagram for realizing third-order quadrature oscillator.

The equation (7) depicts the functions of one current mode inverting lossy integrator and two current mode non-inverting lossless integrators respectively of Fig. 3 as;

$$\frac{I_B}{I_A} = -\frac{a}{s+a}, \quad \frac{I_C}{I_B} = \frac{b}{s}, \quad \frac{I_A}{I_D} = \frac{c}{s} \tag{7}$$

And $$I_D = I_B + I_C = \left(1+\frac{b}{s}\right) I_B \tag{8}$$

From (7) and (8), $$\frac{I_A}{I_D} = \frac{I_A}{\left(1+\frac{b}{s}\right) I_B} = \frac{c}{s} \tag{9}$$

From (7) and (9); $$\frac{I_A}{I_B} = -\frac{s+a}{a} = \frac{c}{s}\left(1+\frac{b}{s}\right) \tag{10}$$

By solving (10), following characteristic equation is obtained:

$$s^3 + as^2 + acs + abc = 0 \tag{11}$$

From (11), frequency of oscillation (FO) is obtained by equating term, $as^2 + abc = 0$, and results as;

$$\text{FO:} \quad \omega_O = \sqrt{bc} \tag{12}$$

Further, by using (12) in the term, $s^3 + acs = 0$ of (11) results the CO as;

$$\text{CO: } a = b \tag{13}$$

Since, (11) represents the characteristic equation of a third-order oscillator, thus the design based on configuration of Fig. 3 will synthesize a third-order quadrature oscillator.

In line with Fig. 3, the proposed QO is designed. It consists of one analog building block (ABB), MDXCCCTA, and three grounded passive capacitors, C_1, C_2, and C_3, as shown in Fig. 4. The terminals Z_{+C} and Z_{+C2} are the Z_+ copies, $Z\text{-}_C$ is the Z- copy, and O_{1+C} is the O_{1+} copy of MDXCCCTA. The V_{OA} and V_{OB} represent the voltages across the capacitors, C_1 and C_2, respectively. Nodes A, B, C, D corresponding to Fig. 3 are mentioned in the proposed circuit of Fig. 4. The parameters a, b and c, of Fig. 3 correspond to the circuit realization terms in proposed oscillator configuration of Fig. 4, as;

$$a = \frac{1}{R_{X+}C_1}, \quad b = \frac{g_{m1}}{C_2} \quad \text{and} \quad c = \frac{1}{R_{X-}C_3} \tag{14}$$

Routine analysis of the circuit in Fig. 4 results in the characteristic equation of a TOQO as:

$$s^3 C_1 C_2 C_3 R_{X+} R_{X-} + s^2 C_2 C_3 R_{X-} + sC_2 + g_{m1} = 0 \tag{15}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

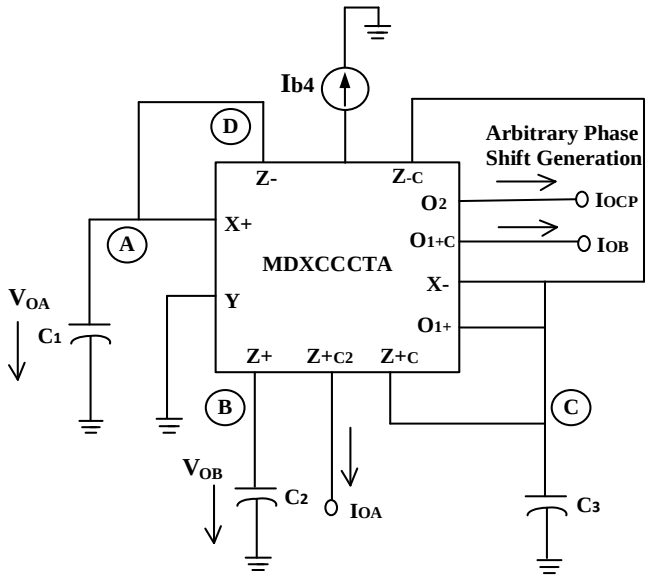


Fig. 4. Proposed TOQO circuit with arbitrary phase shift oscillator.

FO and CO are obtained as follows;

$$\text{FO:} \quad \omega_O = \sqrt{\frac{g_{m1}}{R_{X\text{-}}C_2C_3}} = \sqrt{\frac{(8k_2I_{b2})^{1/2}(k_3I_{b3})^{1/2}}{C_2C_3}} \tag{16}$$

$$\text{CO:} \quad g_{m1}R_{X+} = \frac{C_2}{C_1}, \quad \text{or} \quad \frac{(k_3I_{b3})^{1/2}}{(8k_1I_{b1})^{1/2}} = \frac{C_2}{C_1} \tag{17}$$

It is apparent from expressions (16) and (17) that the tuning of $R_{X\text{-}}$ through I_{b2} can control the FO without affecting the CO; similarly, the CO can be controlled by tuning R_{X+} through I_{b1} and C_1 without affecting FO. Hence, the proposed oscillator provides independent or orthogonal controlling of FO and CO through electronic tuning.

As shown in the configuration of Fig. 4, the two quadrature current outputs, I_{OA} and I_{OB}, and two quadrature voltage outputs, V_{OA} and V_{OB}, are simultaneously obtained. Quadrature currents are obtained at high impedance terminals and are related as;

$$I_{OA} = j\frac{\omega_O C_2 I_{OB}}{g_{m1}} \tag{18}$$

Similarly, quadrature voltage outputs are related as;

$$V_{OA} = -j\omega_O C_2 R_{X+} V_{OB} \tag{19}$$

It is observed from (18) and (19) that amplitudes of I_{OA} and I_{OB} and V_{OA} and V_{OB} can be electronically controlled through tuning of g_{m1} and R_{X+}, respectively.

2) Arbitrary Phase Shift Generator/ Oscillator (APSG/APSO)

The proposed oscillator works as an arbitrary phase shift generator/oscillator (APSO) and generates a signal of various phases when I_{b4} in the circuit of Fig. 4 is connected to a d.c. bias current. In many instrumentation techniques with transducers, an APSO [8] is used to compensate for phase errors. It is also required in phase modulators, mixers, and phased array antennas. The proposed oscillator provides arbitrary phase-shifted signals along with quadrature signals.

The output current, I_{OC} corresponds to I_{OCP} at the O_2 terminal of Fig. 4, when QO works as a phase shift generator and g_{m2P} is transconductance being controlled through I_{b4}. I_{OCP} results in:

$$I_{OCP} = g_{m2P}\{(V_{OB} - V_{OA})\} \tag{20}$$

As V_{OA} and V_{OB} are in quadrature, hence as per (19), these may be written as;

$$V_{OB} = V_S sin(\omega_O t) \quad \text{and} \quad V_{OA} = -V_C cos(\omega_O t) \tag{21}$$

Where V_S and V_C are amplitudes. Hence, from (20) and (21), I_{OCP} results in:

$$I_{OCP} = g_{m2P}\{Asin(\omega_O t + \alpha)\} \tag{22}$$

Where, $A = \sqrt{V_S^2 + V_C^2}$ and

$$\text{phase shift} = \alpha = tan^{-1}\left(\frac{V_C}{V_S}\right) \tag{23}$$

For equal amplitudes, $V_S = V_C = V_M$, as per (22) and (23), I_{OCP} results in a current signal having 45^0 of phase-shifting with respect to V_{OB} and is expressed as;

$$I_{OCP} = g_{m2P}\{\sqrt{2}V_M sin(\omega_O t + 45^0)\} \tag{24}$$

Consequently, from (24), the resultant I_{OCP} is a 45^0 phase-shifted signal with respect to I_{OB} as I_{OB} corresponds to $\sin(\omega_O t)$ and, subsequently, -45^0 with respect to I_{OA} as I_{OA} corresponds to $\cos(\omega_O t)$. Hence, there is the simultaneous availability of three progressively 45^0 phase-shifted current signals as, $I_{OB} \rightarrow \sin(\omega_O t)$, $I_{OCP} \rightarrow \sin(\omega_O t + 45^0)$, and $I_{OA} \rightarrow \sin(\omega_O t + 90^0) \rightarrow \cos(\omega_O t)$, along with two quadrature signals, I_{OA} and I_{OB}.

Further, it is observed from (19) that the relative amplitudes of V_{OB} and V_{OA} can be adjusted by controlling R_{X+} with I_{b1}. If the amplitudes of generated quadrature signals, V_{OA} and V_{OB} are being varied relative to each other by selecting different values of R_{X+} through tuning of I_{b1}, the phase shift (α) of the I_{OCP} will be in the range either from 0 to 45^0 or 45^0 to 90^0 with respect to I_{OB}, depending on $V_S > V_C$ or $V_C > V_S$ as per (23). Simultaneously, g_{m1} is adjusted through I_{b3}, and R_{X-} is adjusted through I_{b2} to maintain the CO and FO constant. Variation in I_{b1} (i.e., R_{X+}) results in a different set of amplitudes of V_{OA} and V_{OB}, and subsequently, the resultant I_{OCP} is a phase shift generated signal with respect to I_{OB}. The phase value of the generated signal (I_{OCP}) is established through (23), and its magnitude is adjusted to the required value through tuning of g_{m2P} (i.e., I_{b4}) as per (22). Therefore, any arbitrary phase shift in the range of 0^0 to 90^0 with respect to I_{OB} and in the range of -90^0 to 0^0 with respect to I_{OA} can be obtained.

C. Amplitude Modulator (AM)

Several instrumentation techniques require AM generation and employ analog multipliers as modulators [4]. However, the use of multipliers makes the system complex. The proposed TOQO circuit works as an AM signal generator from the same circuit configuration of Fig. 4. It is obtained by just connecting the I_{b4} of Fig. 4 to an a.c. signal. The TOQO with AM generation is shown in Fig. 5. The output current (I_{OC}) corresponds to I_{OCM} when QO works as an AM generator. Thereby, the I_{OCM} can be controlled with a.c. signal, I_{b4} is expressed as;

$$I_{OCM} = g_{m2M}(V_{OB} - V_{OA}) \tag{25}$$

The value of transconductance (g_{m2M}) varies with the a.c.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

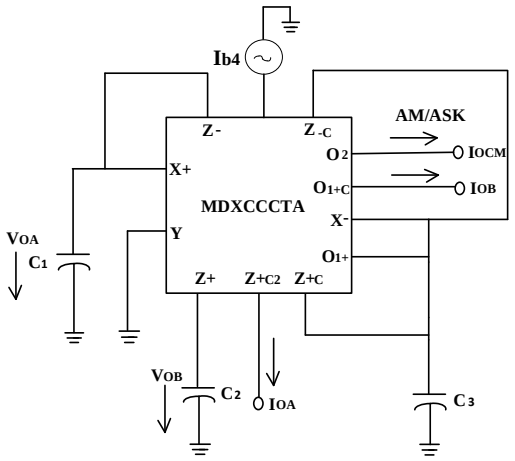


Fig. 5. Proposed TOQO circuit with AM generation.

signal, I_{b4}, without affecting the CO and FO. Hence, if a signal is applied at I_{b4}, it acts as a modulating signal, and as a result, the AM/ASK modulated signals can be obtained from I_{OCM}. Hence, on applying a sinusoidal or triangular as I_{b4}, an amplitude modulated (AM) signal resulted at I_{OCM}, whereas on applying a rectangular signal, an amplitude shift keying (ASK) signal is obtained. Therefore, the proposed TOQO works as an AM signal generator and an APSO from the same circuit topology by altering the connection at I_{b4} from a d.c. to an a.c. current source without requiring any additional component.

III. Automatic Control of CO and FO

In this section, a method is proposed to automatically control the CO and FO of TOQO to stabilize it against variations in relative amplitudes of V_{OA} and V_{OB}. This approach is illustrated in Fig. 6. Firstly, the peaks of generated quadrature signals, V_{OA} and V_{OB}, are detected through a peak detector (Peak Det.), and

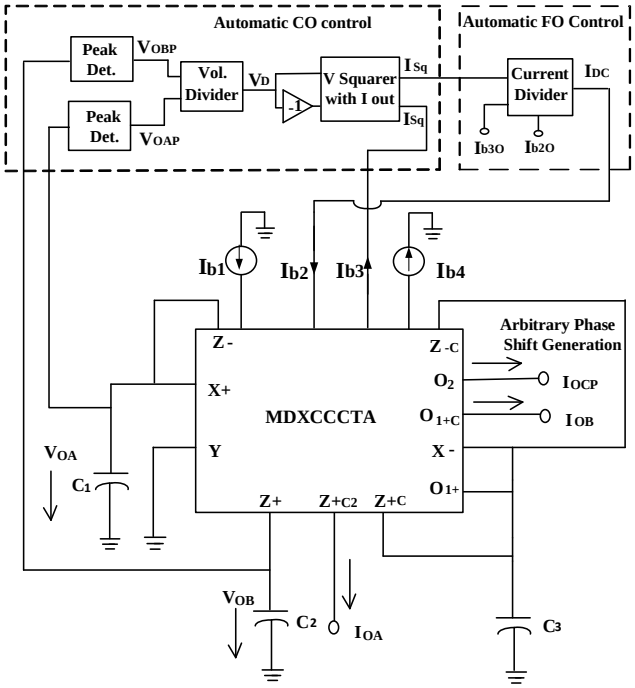


Fig. 6 Automatic CO and FO control in proposed TOQO.

then these two detected (V_{OAP}, V_{OBP}) signals get divided through a voltage divider. The resultant divider voltage signal (V_D) and its inverted signal, $-V_D$, are fed to a voltage (V) squarer with I out, the output of which is a current signal (I_{Sq}) proportional to the square of the input voltage (V_D). This current signal (I_{Sq}) is fed as I_{b3} in the MDXCCCTA, which is an automatically adjusted value of I_{b3} to maintain CO. Furthermore, a replica of the current, I_{Sq}, is fed to a current divider circuit, which generates a current (I_{DC}) in a reverse proportion to I_{Sq}. I_{DC} is fed as I_{b2} in MDXCCCTA and is an automatically adjusted value of required I_{b2} to maintain FO constant.

The governing approach of a proposed automatic control circuit configuration of Fig. 6 is illustrated as follows.

I_{b2O} and I_{b3O} correspond to the values of I_{b2} and I_{b3}, respectively, at which $V_{OAP} = V_{OBP}$ for the circuit's specified operating frequency (say 30.59 MHz).

From (17), the CO is:

$$\frac{\sqrt{k_3 I_{b3}}}{\sqrt{8k_1 I_{bl}}} = \frac{C_2}{C_1} \tag{26}$$

For $\frac{C_2}{C_1} = 1$; $\quad \frac{\sqrt{k_3 I_{b3}}}{\sqrt{8k_1 I_{bl}}} = 1$ or $I_{b3} = \frac{8k_1 I_{b1}}{k_3}$ (27)

As per (19), R_{X+} is expressed as;

$$R_{X+} = j\frac{V_{OA}}{\omega_O C_2 V_{OB}} \tag{28}$$

If, $\omega_O C_2$=constant=G_1, then; (29)

$$R_{X+} \propto \frac{V_{OAP}}{G_1 V_{OBP}} \tag{30}$$

Where, V_{OAP}= peak voltage of V_{OA}, and V_{OBP}= peak voltage of V_{OB} detected by a peak detector.

(30) implies that on varying R_{X+}, proportionally the ratio of peak values of generated quadrature signals ($\frac{V_{OAP}}{V_{OBP}}$) will be changed on keeping G_1 constant.

From (2) and (30), we get $\frac{1}{\sqrt{8k_1 I_{bl}}} \propto \frac{V_{OAP}}{G_1 V_{OBP}}$ (31)

Hence, $\frac{1}{8k_1 I_{bl}} \propto \left(\frac{V_{OAP}}{G_1 V_{OBP}}\right)^2$ or $\frac{1}{I_{bl}} \propto \frac{8k_1}{{G_1}^2}\left(\frac{V_{OAP}}{V_{OBP}}\right)^2$ (32)

If $\frac{8k_1}{{G_1}^2} = G_2$ =constant in (32) then; (33)

$$\frac{1}{I_{bl}} \propto G_2\left(\frac{V_{OAP}}{V_{OBP}}\right)^2 \text{ or, } I_{bl} = \frac{1}{G_2}\left(\frac{V_{OBP}}{V_{OAP}}\right)^2 \tag{34}$$

On considering, $\left(\frac{V_{OBP}}{V_{OAP}}\right)^2 = M$ (35)

From (34) results, $I_{bl} = \frac{M}{G_2}$ (36)

It is inferred from (36) that on keeping G_2 constant, variation in I_{b1} is directly proportional to a change in the value of M.

From (27) and (36), we get

$$I_{b3} = \frac{8k_1}{k_3 G_2} M \tag{37}$$

If, $\frac{8k_1}{k_3 G_2} = G_3$, then (38)

$$I_{b3} = G_3 M = I_{b3O} M \tag{39}$$

Therefore, it may be noticed, from (36) and (39), that after each variation in I_{b1}, M will be changed proportionately; as a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

consequence, I_{b3} will also be changed, and subsequently, CO will automatically be maintained constant as per (26).

The peak voltages, $V_1=V_{OAP}$ and $V_2=V_{OBP}$ obtained through peak detector [45] are subsequently divided by a simple OTA-based Vol. Divider [46] and results in:

$$V_D = K_V \frac{V_{OBP}}{V_{OAP}}; \quad (40)$$

Where K_V is the proportionality constant [46].

The results in V_D and inverted V_D (through an inverter) are fed to a simple CCII-based 'V Squarer with I out' circuit as shown in Fig. 7. It is based on [47] with the addition of a z- terminal. The output of it is a current signal (I_{Sq}) proportional to the square of V_D and is expressed as:

$$I_{Sq} = K_{Sq} V_D^2 = K_{Sq}\left(K_V \frac{V_{OBP}}{V_{OAP}}\right)^2 = K_{Sq} K_V^2 \left(\frac{V_{OBP}}{V_{OAP}}\right)^2 \quad (41)$$

$K_{Sq} = \mu_n C_{OX}\left(\frac{W}{L}\right)_{NMOS}$ is of matched NMOS of Fig. 7. This current I_{Sq}, as shown in Fig. 6, is used as I_{b3}.

Thus, (39) and (41) results in:

$$I_{Sq} = I_{b3} = K_{Sq} K_V^2 M = I_{b3O} M \quad (42)$$

Hence, on varying relative amplitudes of V_{OAP} and V_{OBP} with I_{b1}, the M value will change, and as a consequence, $I_{Sq} = I_{b3}$ changes with factor M. Thus, CO is automatically maintained as per (26).

Furthermore, to maintain the FO constant, simply a current divider [48] is employed in the proposed automatic control system. As shown in Fig. 6, a copy of obtained I_{Sq} is fed through the z+ terminal of "V Squarer with I out" as shown in Fig. 7 to the input of the current divider and output I_{DC} results as [48];

$$I_{DC} = \frac{I_X I_Y}{I_w} \quad (43)$$

If, $I_W = I_{Sq} = I_{b3} = I_{b3O}\left(\frac{V_{OBP}}{V_{OAP}}\right)^2$, $I_X = I_{b3O}$, $I_Y = I_{b2O}$, then from (35);

$$I_{b2} = I_{DC} = \frac{I_{b2O}}{\left(\frac{V_{OBP}}{V_{OAP}}\right)^2} = \frac{I_{b2O}}{M} \quad (44)$$

I_{DC} is connected as I_{b2} in the proposed circuit, hence $I_{DC} = I_{b2}$. From (44) and (42), we get

$$I_{b2} = \frac{I_{b2O}}{M} \text{ and } I_{b3} = M I_{b3O} \quad (45)$$

Moreover, from (16); FO: $\omega_O = \sqrt{\frac{\sqrt{8k_2 I_{b2}}\sqrt{k_3 I_{b3}}}{C_2 C_3}}$ and on keeping C_2C_3=constant, ω_O results in:

$$\omega_O \propto \sqrt{\sqrt{8k_2 I_{b2}}\sqrt{k_3 I_{b3}}} \propto \sqrt{\sqrt{\frac{8k_2 I_{b2O}}{M}}\sqrt{k_3 M I_{b3O}}} \quad (46)$$

Therefore, for each variation of I_{b1}, the value of I_{b3} varies with a factor of M, and I_{b2} varies with a factor of 1/M as per (45). Thus, the FO remains unaffected as per (46).

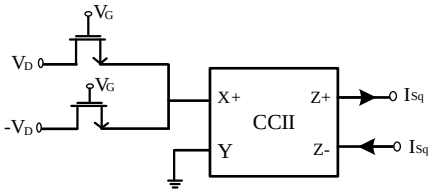


Fig. 7 V Squarer with I out circuit.

Therefore, it is confirmed that ω_O will be maintained to a fixed value through the automatic adjustment of I_{b2}, and CO will automatically be adjusted through I_{b3} as per (42). The proposed stabilization approach results as a new technique to automatically control the CO and FO simultaneously.

IV. Nonideal Analysis

The frequency performance of the oscillator circuit may deviate from the ideal one due to the non-ideality of MDXCCCTA. The non-ideality effects may be categorized into two groups. The first group is due to tracking errors of internal current and voltage transfers of MDXCCCTA. The second group is from the parasitics of MDXCCCTA comprising resistances and capacitances. The effects of the non-ideality of the MDXCCCTA are discussed as follows:

A. Non-ideality Effect due to Current and Voltage Transfer Gains

Considering the non-ideality of current and voltage transfer gains for the oscillator, the port relationships (1) may be modified as follows;

$$\begin{pmatrix} \boldsymbol{I_Y} \\ \boldsymbol{V_{X+}} \\ \boldsymbol{V_{X-}} \\ \boldsymbol{I_{Z+}} \\ \boldsymbol{I_{Z-}} \\ \boldsymbol{I_{O1+}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{\Phi_1} & \boldsymbol{R_{X+}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{-\Phi_2} & \boldsymbol{R_{X-}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\alpha_1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\beta_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{+\gamma_1 g_{m1}} \end{pmatrix} \begin{pmatrix} \boldsymbol{V_Y} \\ \boldsymbol{I_{X+}} \\ \boldsymbol{I_{X-}} \\ \boldsymbol{V_{Z+}} \end{pmatrix} \quad (47)$$

and $\boldsymbol{I_{Z+C} = \alpha_2 I_{x+}}$, $\boldsymbol{I_{Z\text{-}C} = \beta_2 I_{X-}}$, (48)

Where $\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_2}$, $\boldsymbol{\beta_1}$, and $\boldsymbol{\beta_2}$ are the non-unity current transfer gains, $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ are non-unity voltage transfer gains. Similarly, $\boldsymbol{\gamma_1}$ is the non-ideal transconductance gain. Ideally, all these transfer gains are unity. Considering non-ideal relationships (47) and (48), the characteristic equation (15) of the oscillator is modified as follows;

$$\boldsymbol{s^3 C_1 C_2 C_3 R_{X+} R_{X-} + s^2\{C_2 C_3 R_{X-} + (1-\beta_2) C_1 C_2 R_{X+}\} + s}$$

$$\boldsymbol{\{\beta_1 \alpha_2 C_2 + (1-\beta_2) C_2\} + \beta_1 g_{m1} \gamma_1 \alpha_1 = 0} \quad (49)$$

The FO of (16) and CO of (17) from (49) are modified as follows;

$$\text{FO:} \quad \boldsymbol{\omega_O = \sqrt{\frac{\alpha_1 \beta_1 \gamma_1 g_{m1}}{R_{X-} C_2 C_3 + (1-\beta_2) C_1 C_2 R_{X+}}}} \quad (50)$$

$$\text{CO:} \quad \boldsymbol{\frac{\alpha_1 \gamma_1 g_{m1}}{\alpha_2 + 1 - \beta_2} = \frac{R_{X-} C_2 C_3 + (1-\beta_2) C_1 C_2 R_{X+}}{C_1 C_3 R_{X+} R_{X-}}} \quad (51)$$

It is observed that FO and CO are affected by transfer gains, whereas it may be noted that, for ideal values of transfer gains, i.e. ($\boldsymbol{\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = \gamma_1 = 1}$), (49)-(51) result in original expressions as of (15), (16) and (17).

B. Non-ideality Effect due to Parasitic Components

The second group of non-ideality comes from parasitic resistances and capacitances of MDXCCCTA. The QO using the parasitic model of MDXCCCTA is shown in Fig. 8.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

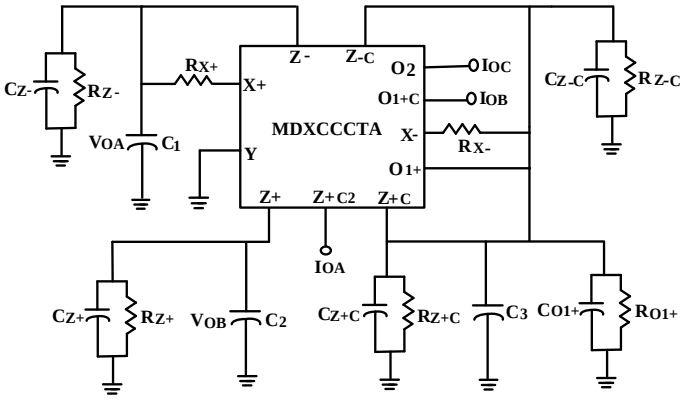


Fig. 8 Non-ideal equivalent of MDXCCCTA oscillator circuit.

Parasitic resistances and capacitances appear in parallel at terminals Z_+, Z-, Z_{+C}, $Z\text{-}_C$, and O_{1+} and intrinsic resistances in series at terminals X_+ and X-. Parasitic capacitances (C_{1P}, C_{2P}, C_{3P}) and resistances (R_{1P}, R_{2P}, R_{3P}) are associated in parallel to C_1, C_2, and C_3 in circuit of Fig. 4. Equivalent capacitances (C_{eq1}, C_{eq2}, C_{eq3}) at different terminals are expressed as;

$$C_{eq1} = C_1 // C_{1P},\ C_{eq2} = C_2 // C_{2P}, \text{ and } C_{eq3} = C_3 // C_{3P} \quad (52)$$

$$\text{Where, } C_{1P} = C_{Z-}, C_{2P} = C_{Z+}, C_{3P} = C_{Z+C} // C_{Z-C} // C_{O1+} \quad (53)$$

Similarly, equivalent parasitic resistances are as follows;

$$R_{1P} = R_{Z-},\ R_{2P} = R_{Z+}, \text{ and } R_{3P} = R_{Z+C} // R_{Z-C} // R_{O1+}. \quad (54)$$

On considering the values from (52)-(54), the modified characteristic equation of oscillator results as;

$$s^3 + s^2\left(\frac{1}{h_1}+\frac{1}{h_2}+\frac{1}{h_3}+\frac{1}{n_1}\right) + s\left(\frac{1}{h_2h_3}+\frac{1}{h_1h_2}+\frac{1}{h_1h_3}+\frac{1}{n_2}+\frac{1}{n_3}+\frac{1}{n_4}\right)+\frac{1}{n_9}\left(n_8+\frac{1}{n_5}+\frac{1}{n_6}+\frac{1}{n_7}\right)=0 \quad (55)$$

Where, $h_1 = C_{eq1}R_{1p}$, $h_2 = C_{eq2}R_{2p}$, $h_3 = C_{eq3}R_{3p}$;

$n_1 = C_{eq1}R_{X+}$, $n_2 = C_{eq1}C_{eq2}R_{3p}R_{X+}$, $n_3 = C_{eq1}C_{eq2}R_{2p}R_{X+}$,

$n_4 = C_{eq1}C_{eq3}R_{X+}R_{X-}$, $n_5 = R_{1p}R_{2p}R_{3p}$, $n_6 = R_{2p}R_{3p}R_{X+}$,

$n_7 = R_{2p}R_{X+}R_{X-}$, $n_8 = \frac{g_{m1}}{R_{X+}R_{X-}}$, $n_9 = C_{eq1}C_{eq2}C_{eq3}$ (56)

In order to minimize the effect of parasitic components of MDXCCCTA, the following relations should be satisfied:

$$\frac{1}{|(s(C_{eq1})|} \ll R_{1P},\quad \frac{1}{|(s(C_{eq2})|} \ll R_{2P},\quad \frac{1}{|(s(C_{eq3})|} \ll R_{3P}\ ; \text{ or } \quad (57)$$

$$\frac{1}{R_{1P}(C_{eq1})} \ll |s|,\quad \frac{1}{R_{2P}(C_{eq2})} \ll |s|,\quad \frac{1}{R_{3P}(C_{eq3})} \ll |s| \quad (58)$$

$$\text{Let, } \frac{1}{R_{1P}(C_{eq1})} = \omega_{p1},\ \frac{1}{R_{2P}(C_{eq2})} = \omega_{p2},\ \frac{1}{R_{3P}(C_{eq3})} = \omega_{p3} \quad (59)$$

As, $s = j\omega_O$, hence (58) and (59) imply that;

$$\text{Operating frequency } (\omega_O) \gg \omega_{p1}, \omega_{p2}, \omega_{p3} \quad (60)$$

Equation (60) depicts that, to minimize the effect of parasitics, the operating oscillation frequency (ω_O) must be the highest of [ω_{p1}, ω_{p2}, ω_{p3}]. Moreover, as the parasitic capacitances are in parallel to external capacitors, the effect can be ignored if $C_1 >> (C_{1P})$, $C_2 >> (C_{2P})$, and $C_3 >> (C_{3P})$.

Furthermore, the simplified expression for oscillation frequency can be obtained by considering R_{1P}, R_{2P}, and R_{3P} are high resistance values, and C_{1P}, C_{2P}, C_{3P} are low capacitance values, which is valid for MDXCCCTA. Under these conditions coefficient of s^2, s, and constant in (55) except terms $1/n_1$, $1/n_4$, and n_8/n_9 result insignificant and can be neglected. Therefore, characteristic equation (55) gets simplified as;

$$s^3 + \frac{s^2}{C_{eq1}R_{X+}} + \frac{s}{C_{eq1}C_{eq3}R_{X+}R_{X-}} + \frac{g_{m1}}{C_{eq1}C_{eq2}C_{eq3}R_{X+}R_{X-}} = 0 \quad (61)$$

The FO and CO from (61) are as follows;

$$\text{FO: } \boldsymbol{\omega_O} = \sqrt{\frac{g_{m1}}{R_{X-}C_{eq2}C_{eq3}}} \quad (62)$$

$$\text{CO: } g_{m1}R_{X+} = \frac{C_{eq2}}{C_{eq1}} \quad (63)$$

$$\text{If } C_1 >> C_{1P}, C_2 >> C_{2P} \text{ and } C_3 >> C_{3P}, \text{ then} \quad (64)$$

$$C_{eq1} \to C_1, C_{eq2} \to C_2 \text{ and } C_{eq3} \to C_3. \quad (65)$$

It is observed that, on considering condition (65) in (61)-(63), the expressions (61)-(63) result in the ideal form of (15), (16), and (17), respectively.

V. Simulation Results And Discussion

The developed TOQO circuit is simulated using 180nm CMOS process technology in Cadence Virtuoso with model parameters as; C_{OX}=8.422fF/μm^2, $\mu_{(N)}$=259.53cm^2/V-s, $\mu_{(P)}$=109.97cm^2/V-s, and t_{OX}=4.1nm. The results are obtained with parameter values as; supply voltage, $V_{DD} = -V_{SS}$ =1V, bias currents, $I_{b1} = I_{b2}$ =10μA, I_{b3} =60μA, $C_1 = C_2 = C_3$=1.3pF, R_{X+} = 4kΩ, R_{X-} =4kΩ, and g_{m1}=250μA/V. Selected aspect ratios of transistors in MDXCCCTA are shown in Table I.

TABLE I
ASPECT RATIOS OF TRANSISTORS IN MDXCCCTA

Transistor	Width (μm)	Length(μm)
M_1–M_6, M_{29}–M_{36}	2.6	0.36
M_7–M_{13}, M_{17} – M_{28}, M_{41}, M_{42}	1.3	0.36
M_{14}, M_{15}, M_{16}	8, 2, 8	0.36
M_{37}- M_{40}, M_{45}, M_{46}	4	0.36
M_{43}, M_{44}	1.8	0.36
M_{47}, M_{48}	0.5	0.36

Response of the quadrature voltage signals V_{OA} and V_{OB} are shown in Fig. 9. Similarly, the quadrature current signals I_{OA} and I_{OB} are shown in Fig. 10. It can be observed that the voltage and current outputs are 90^0 phase-shifted with each other.

Further, variation in intrinsic resistance, R_{X-} on varying bias current, I_{b2} is shown in Fig. 11. Variation in the frequency of oscillation (f_O) on varying bias current, I_{b2} is shown in Fig. 12, depicting f_O is varied on the tuning of I_{b2} over a significant range as per (16), and CO is maintained independently through tuning of bias current, I_{b1} as per (17). It must be noted that Fig. 12 depicts the variation range of f_O through electronic tuning only; however, the lower limit of f_O can further be lowered up to a few kHz on selecting high values of C_1 and C_2 (in range of 1nF). Therefore, the circuit possesses a wide tunable range of oscillation frequency and is suitable for various low to high-frequency applications in instrumentation techniques. The frequency spectrum corresponding to oscillation across V_{OA} is shown in Fig. 13. Fig. 13 illustrates that for the selected set of parameter values as have been

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

mentioned in the beginning of this section (i.e. $C_1=C_2=C_3=1.3$pF, $R_{X+}=R_{X-}=4k\Omega$, $g_{m1}=250\mu A/V$), the f_O of the voltages (V_{OA}, V_{OB}) and currents (I_{OA}, I_{OB}) corresponds to 30.59MHz and is one of the maximum reported values in the available literature. Simulated f_O closely follows the theoretically designed value, 30.6MHz.

Further, the percentage total harmonic distortion (%THD) of voltage and current signals are shown in Fig. 14(a) and (b), respectively, and the depicted values are less than 0.8% for V_{OA}, V_{OB} and less than 0.84% for I_{OA}, I_{OB}. Fig. 15 shows the phase error, i.e., deviation from quadrature phase value 90^0, calculated from Fourier analysis simulation values. Phase error for V_{OA} and V_{OB} is less than 0.2% and for I_{OA} and I_{OB} is less than 0.3%, which is considered very small. Moreover, the phase noise analysis has been done and the simulated phase noise (dBc) versus frequency responses for V_{OA} and V_{OB} are shown in Fig. 16. It is observed that phase noise is of low-level and in the operating range of proposed circuit (822k-52.94MHz), it is between -180dBc to -240dBc, which is significantly very low.

Furthermore, to verify the robustness of the proposed oscillator, the Monte Carlo (MC) simulation for variation in the f_O of V_{OA} for 200 runs is shown in Fig. 17. Statistical distribution of MC simulation outcome shows oscillator is vastly robust against tolerance in passive components as f_O remains thoroughly close to a standard value, 30.6MHz. Moreover, the circuit's layout has been designed to validate the proposed TOQO for physical verification or fabrication, as shown in Fig. 18, which consumes an area of $7810\mu m^2$. Post layout time series simulation results for signals, I_{OA}, and I_{OB} are shown in Fig. 19, depicting responses are close to the original form except for the inclusion of slight delay caused by parasitic components and; subsequently, the frequency of oscillation deviates slightly. Deviation from 90^0 is 0.16^0 i.e. phase error of 0.18% is observed in the response, which is very small.

To illustrate the working of the circuit as an arbitrary phase shift oscillator (APSO), the simulation results are depicted in Fig. 20-23. The plot for I_{OCP} has 45^0 phase shifts with respect to I_{OB} and - 45^0 with respect to I_{OA} for equal amplitudes of V_{OA}, and V_{OB} is shown in Fig. 20 for I_{b4}= 80μA. The phase shift of 30^0 between I_{OB} and I_{OCP} and subsequently -60^0 between I_{OA} and I_{OCP} are obtained for I_{b1}=30μA as depicted in Fig. 21 with I_{b4}= 42μA. Similarly, the phase shift of 60^0 between I_{OB} and I_{OCP} and -30^0 between I_{OA} and I_{OCP} are obtained for I_{b1}=3μA as depicted in Fig. 22 with I_{b4}= 42μA. Moreover, the continuous variation in phase between I_{OB} and I_{OCP} and between I_{OA} and I_{OCP} with variation in I_{b1} is shown in Fig. 23. Fig. 23 depicts that phase shifting between I_{OB} and I_{OCP} is in the range of 0 to 90^0, whereas phase shift between I_{OA} and I_{OCP} is in -90^0 to 0. It verifies the application of the circuit of Fig. 4 as an APSO.

Moreover, simulated responses are shown in Fig. 24(a)-(c) to validate the proposed circuit as AM generator. If a modulating signal is applied as I_{b4}, the AM/ASK signals can be obtained as I_{OCM} as per (25). In verifying this phenomenon, a sinusoidal current signal, I_{b4} of frequency (f_m) of 3 MHz and peak current of 70 μA is applied; the obtained AM response signal, I_{OCM}, is shown in Fig. 24(a). Further, on applying a triangular signal as I_{b4}, the obtained response of AM signal, I_{OCM}, is shown in Fig. 24(b). Similarly, on applying a rectangular signal as I_{b4}, the obtained response I_{OCM} of modulated signal (ASK) is shown in Fig. 24(c). The modulation index of AM signal in Fig. 24(a) results as 0.74, and the percentage is 74%. The modulation index result of Fig. 24(b) is 0.68, i.e., 68%, and the ASK result in Fig. 24(c) is 0.76, i.e., 76%. The power spectrum plot for AM signal of Fig. 24(a) has been obtained and is shown in Fig. 25 for f_m=3 MHz and $f_c=f_o$=30.59 MHz. It is depicted that second-order intermodulation products (IM2) are obtained at $f_O\pm f_m$, i.e., at 33.59 MHz and 27.59 MHz. Further, it is observed that third-order intermodulation products (IM3) are obtained at $f_O\pm 2f_m$, i.e., at 36.59 MHz and 24.59 MHz, and corresponding intermodulation distortion (IMD3) are -58 dBc and -57 dBc, respectively. It may be observed that intermodulation distortions are of a considerably low level. The power consumption of the proposed TOQO circuit is 1.08 mW.

Moreover, to verify the workability of the proposed automatic CO and FO control system depicted in Fig. 6, the simulated results are presented in Table III and compared with the results of the non-automatic system in Fig. 4. Circuit has been simulated with parameters as; K_V=1V, $K_{Sq}=60\mu A/V^2$ from (40) and (41) respectively by keeping $\mu_nCox=160\mu A/V^2$, W/L=1μ/2.5μ of NMOS of Fig. 7. Moreover, initial values of bias currents are set as, I_{b1O}=10μA and I_{b3O}= 60μA from (45) and M vary according to change in $(V_{OA}/V_{OB})^2$, which varies on the tuning of I_{b1} as per (36). Table III shows that the value of I_{b1} has been varied from 2 μA to 60 μA, and the proposed automatic system results closely match the non-automatic or manually adjusted system results. Automatically obtained values of I_{b2} and I_{b3} are in good agreement with manually adjusted values to maintain the CO and FO. Furthermore, frequency deviation and phase shift deviation between automatic and non-automatic circuits are less than 1% for each varied value of I_{b1}.

Moreover, to check the performance of circuit under extreme conditions, the PVT (process, voltage, temperature) analysis has been done. The process corner simulation for conditions FF (fast NMOS, fast PMOS), FS (fast NMOS, slow PMOS), SF (slow NMOS, fast PMOS), SS (slow NMOS, slow PMOS) and NN (nominal) has been done and the obtained time domain and frequency spectrum responses for V_{OA} are shown in Fig. 26(a) and 26(d) respectively. Further, responses corresponding to temperature and voltage supply variation are shown in Fig. 26(b) and Fig. 26(c) respectively. The effect of PVT corners on frequency (f_o) and amplitude (A_m) is summarized in Table II. It may be noticed that deviation in response from nominal values (30.59 MHz, 0.3V) is within 0.8%, which is considerably small and frequency spectrum of Fig. 26(d) also verifies the process corner response values resulting small deviations, mentioned in Table II.

TABLE II
PVT CORNERS EFFECT

	Process Corner					**Temperature**			**Voltage supply**		
	FF	**FS**	**NN**	**SF**	**SS**	**-40⁰**	**27⁰**	**80⁰**	**1.2**	**1**	**0.8**
f_O(MHz)	30.7	30.68	30.59	30.51	30.4	30.69	30.59	30.53	30.61	30.59	30.58
A_m(V)	0.3018	0.301	0.3	0.299	0.298	0.3013	0.3	0.299	0.3012	0.3	0.2994
% deviation;											
f_O	0.36	0.29	0	0.26	0.62	0.32	0	0.19	0.13	0	0.03
A_m	0.6	0.33	0	0.33	0.66	0.43	0	0.33	0.4	0	0.2

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

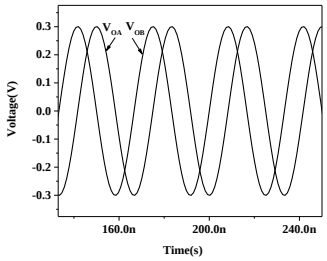


Fig. 9. Quadrature voltage outputs, V_{OA} and V_{OB}.

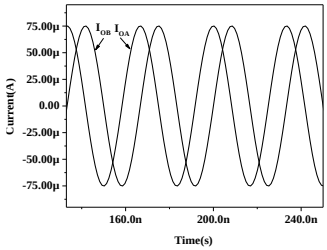


Fig. 10. Quadrature current outputs, I_{OA} and I_{OB}.

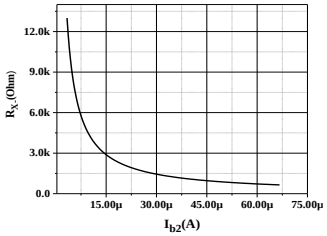


Fig. 11. Variation in intrinsic resistance (R_{X-}) with bias current, I_{b2}.

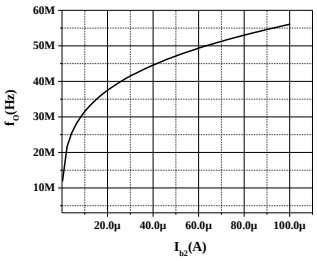


Fig. 12. Variation in frequency of oscillation (f_O) with bias current, I_{b2}.

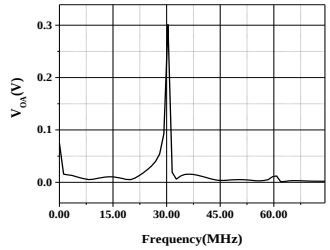


Fig. 13. Frequency spectrum for V_{OA}.

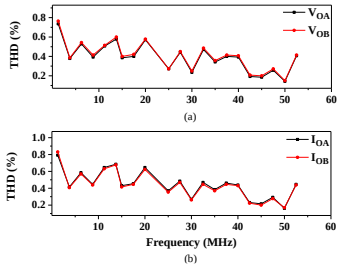


Fig. 14. %THD vs. frequency responses for (a) V_{OA} and V_{OB} (b) I_{OA} and I_{OB}.

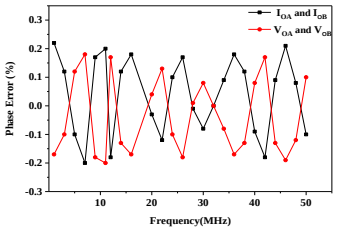


Fig. 15. Phase error vs. frequency between V_{OA} & V_{OB} and I_{OA} & I_{OB}.

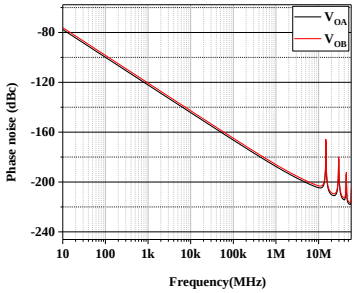
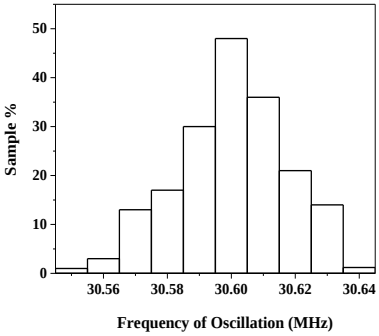


Fig. 16. Phase noise response for quadrature voltages, V_{OA} and V_{OB}.



n samples =200	mean= 30.6MHz	median= 30.595MHz
n division = 10	min = 30.546044MHz	max=30.6432MHz
σ = 0.10048MHz	10%ile = 30518044	90%ile =30632248

Fig. 17. Monte-Carlo simulation for frequency of oscillation of V_{OA}.

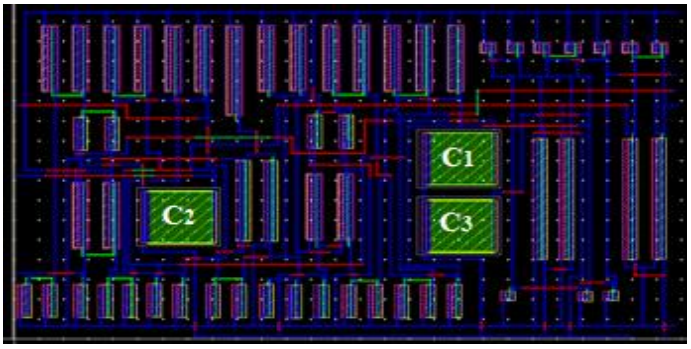


Fig. 18. Layout of proposed TOQO circuit.

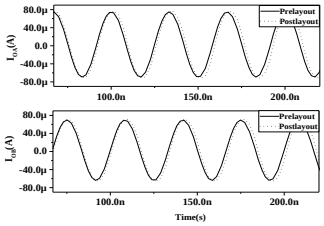


Fig. 19. Post layout responses for I_{OA} and I_{OB}.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

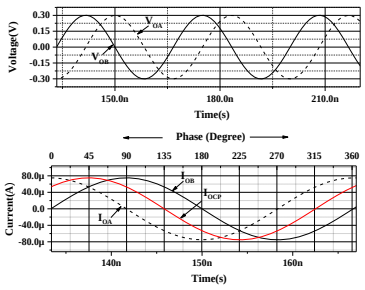


Fig. 20. Plot for I_{OCP} having 45^0 and -45^0 of phase-shifting with respect to I_{OB} and I_{OA}, respectively, for $V_{OA} = V_{OB}$.

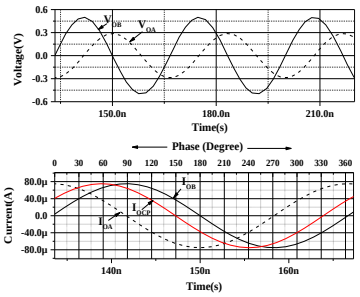


Fig. 21. Plot for I_{OCP} having 30^0 phase-shifting with respect to I_{OB} and -60^0 with respect to I_{OA} for $V_{OB}>V_{OA}$.

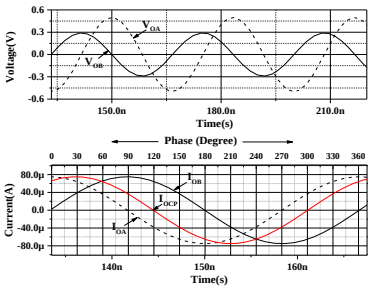


Fig. 22. Plot for I_{OCP} having 60^0 phase-shifting with respect to I_{OB} and -30^0 with respect to I_{OA} for $V_{OA}>V_{OB}$.

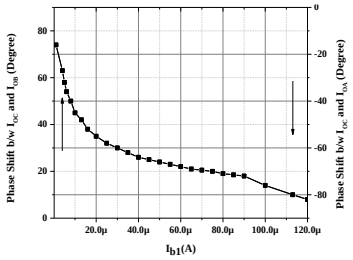
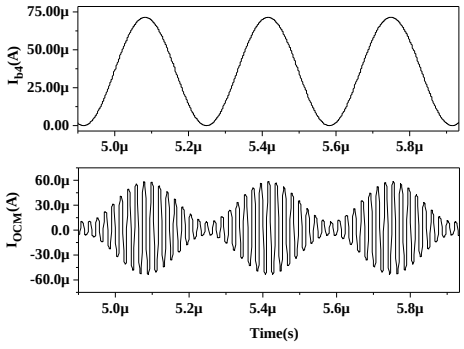
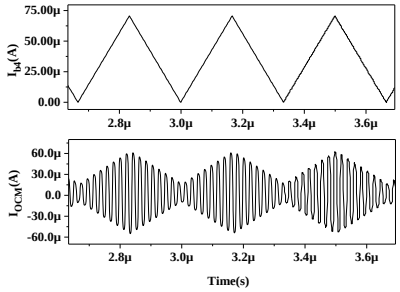


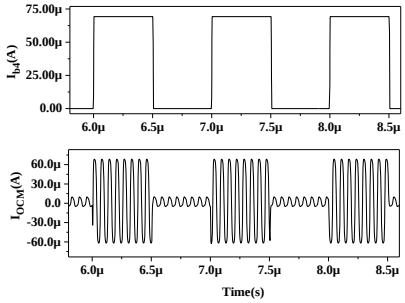
Fig. 23. Plot for phase-shifting between I_{OCP} and I_{OB} in the range of 0 to 90^0 and between I_{OCP} and I_{OA} in the range of -90^0 to 0 with variation in I_{b1}.



(a)



(b)



(c)

Fig. 24. Modulated signals at I_{OCM} for modulating signal, I_{b4} as; (a) sinusoidal, (b) triangular, (c) rectangular for f_c=30.59 MHz and f_m=3 MHz.

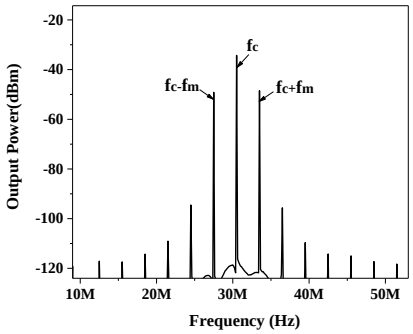


Fig. 25. Power spectrum of AM signal showing intermodulation distortions.

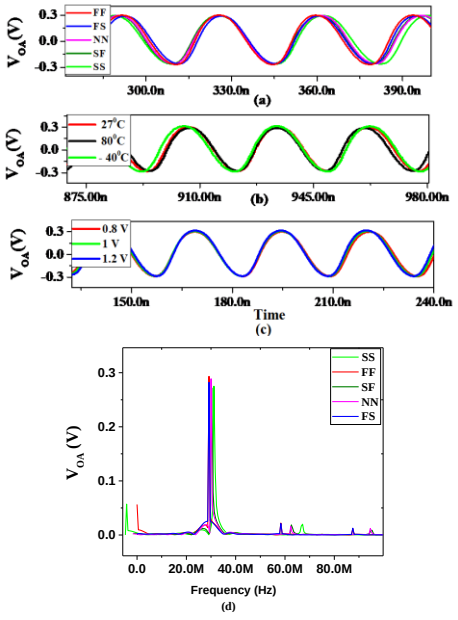


Fig. 26. PVT corner analysis in time domain and frequency spectrum responses for V_{OA}; time domain responses; (a) process corners (b) temperature (c) voltage supply; (d) frequency spectrum for process corners (FF, SF, FS, NN, SS).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

TABLE III
COMPARISON BETWEEN RESULTS OF NON-AUTOMATIC AND AUTOMATIC CO AND FO CONTROL

Manual adjustment of I_{b3} and I_{b2} to control CO and FO					Automatic adjustment of I_{b3} and I_{b2} to control CO and FO					
I_{b1} (μA)	I_{b3} (μA)	I_{b2} (μA)	f_O after variation in I_{b2} (MHz)	Phase shift value (degree)	I_{b3} (μA)	I_{b2} (μA)	f_O after variation in I_{b2} (MHz)	Phase shift value (degree)	% f_O deviation	% Phase shift deviation
2	12	48	30.54	65.9	13	49.52	30.14	65.84	1.31	0.09
5	30	20	30.56	54.8	31	21.64	30.38	54.76	0.58	0.07
8	48	13	30.58	48.2	48.9	13.08	30.28	48.01	0.98	0.39
10	60	10	30.58	45	58	9	30.68	44.84	0.32	0.35
15	90	8	30.59	39.4	89	7.5	30.84	39.31	0.81	0.22
20	120	5	30.59	35.3	118	4.98	30.88	35.18	0.94	0.34
25	150	4	30.6	32.3	148	3.84	30.69	32.48	0.29	0.55
30	180	3	30.61	30	178	2.8	30.78	30.28	0.55	0.93
60	360	1.7	30.84	22	358	1.6	30.98	21.89	0.45	0.50

VI. Experimental Verification

As monolithic implementation of MDXCCCTA is not available, the possible experimental realization of TOQO using commercially available ICs, AD844 and CA3080 is shown in Fig. 27, followed by the assembled circuit on a breadboard is given in Fig. 28. This circuit can broadly verify the functionality; however, detail parameters cannot be comparable with monolithic implementation. Supply voltage is set as 7V, with selected capacitances as, $C_1=C_2=C_3$ =100pF, and resistances, $R_1=R_7$ = $R_2=R_3=R_4=R_5=R_6=R_8=1k\Omega$. The target frequency for experimentation of Fig. 27 was set to 2 MHz. Fig. 29 shows the experimental setup. The obtained quadrature voltage outputs, V_{OA} and V_{OB}, are shown in Fig. 30, with a peak-to-peak value of 180mV. Quadrature current outputs, I_{OA} and I_{OB}, are obtained on connecting 1.5kΩ resistors at Z_{+C2} and O_{1+C} terminals of Fig. 27, and responses are shown in Fig. 31. Further, the AM and ASK responses are shown in Fig. 32 for the sinusoidal, triangular, and square wave modulating signals of 100 kHz applied as I_{b4}. It may be noted that the response of Fig. 32(a) and 32(b) are AM signals, and Fig. 32(c) is an ASK-modulated signal. Experimental results for arbitrary phase shift generation between I_{OCP} and I_{OB} are shown in Fig. 33 (a)-(c). The phase shift of 45^0 obtained for $V_{OA}=V_{OB}$ is shown in Fig. 33(a). The phase shift of 30^0 for $V_{OA}<V_{OB}$ obtained at R_1=300 Ω is shown in Fig. 33(b), whereas the phase shift of 60^0 for $V_{OA}>V_{OB}$ obtained at R_1=700 Ω is shown in Fig. 33(c).

Furthermore, the AD844 and CA3080 macro model-based IC simulation has been done in OrCAD PSpice for the similar circuit configuration shown in Fig. 27 with $C_1=C_2=C_3$ =100pF, resistances, $R_1=R_7=R_2=R_3=R_4=R_5=R_6=R_8=1k\Omega$ and target frequency of 2 MHz. The simulated results have shown good agreement with the experimentally obtained results. The comparison of IC AD844, CA3080-based PSpice macro model simulation, post-layout simulation of monolithic circuit, and experimentally obtained results for various performance parameters has been carried out and shown in Table IV. It can be inferred that % THD, phase error, and sensitivity of IC-based simulation and experimental results closely follow the post-layout simulated results at a specified frequency of 1.84 MHz. However, IC CA3080 has a limitation in bandwidth, as it can operate up to 2 MHz only; hence the maximum experimental frequency of operation and frequency tunable range of the proposed circuit are limited in IC (CA3080 and AD844) based simulation and experimental results. The voltage supply and power consumption are high for discrete IC-based simulation and experimental results.

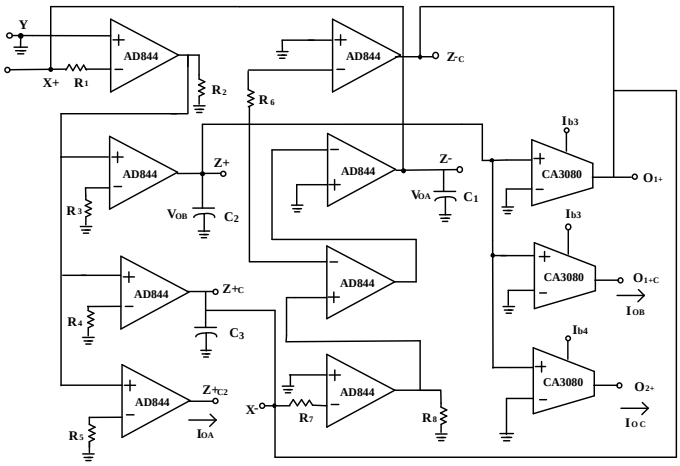


Fig. 27. Implementation of TOQO using commercially available ICs AD844 and CA3080.

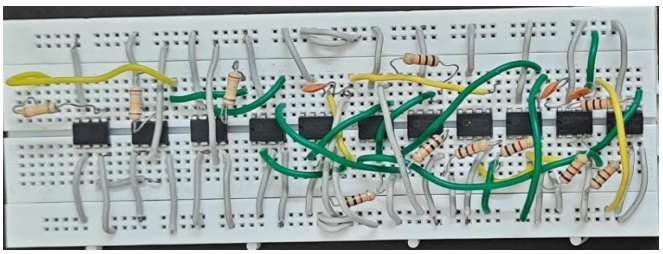


Fig. 28. Prototype of TOQO using AD844 and CA3080.

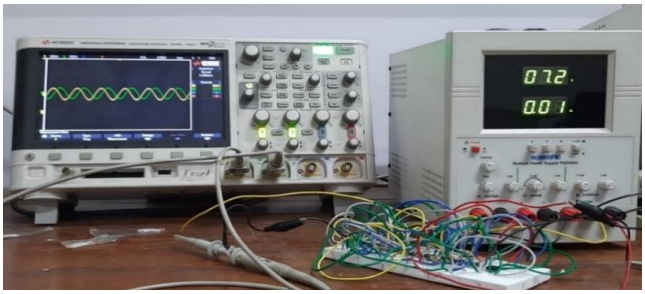


Fig. 29. Experimental setup.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

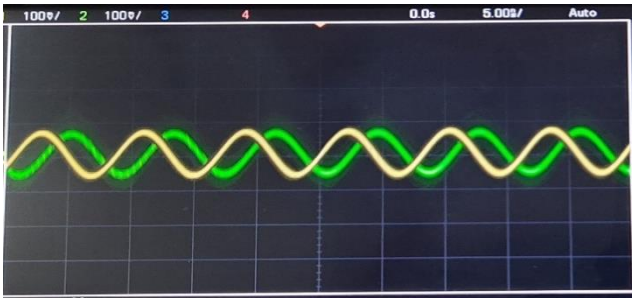


Fig. 30. Quadrature voltage outputs, V_{OA} (green) and V_{OB} (yellow).

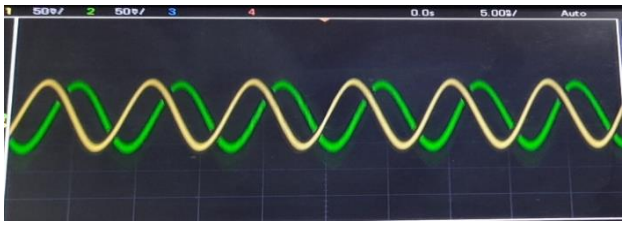
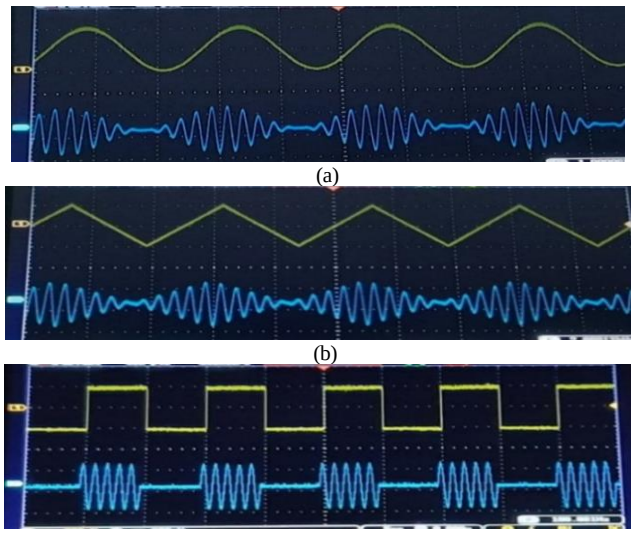
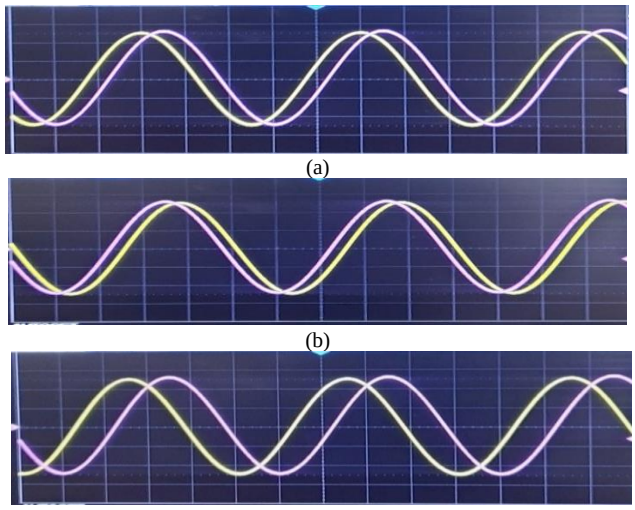


Fig. 31. Quadrature current outputs, I_{OA} (yellow) and I_{OB} (green).



(a) (b) (c)

Fig. 32. Responses of modulated outputs, I_{OC} for modulating signal, I_{b4} of (a) Sinusoidal (b) Triangular (c) Square.



(a) (b) (c)

Fig. 33. Responses for arbitrary phase shift between I_{OCP} and I_{OB} ;(a) 45^0 for $V_{OA}=V_{OB}$, (b) 30^0 for $V_{OA}<V_{OB}$, (c) 60^0 for $V_{OA}>V_{OB}$.

TABLE IV
COMPARISON OF SIMULATION AND EXPERIMENTAL RESULTS

Performance Parameter	Monolithic Post layout simulation	Discrete IC based PSpice simulation	Experimental using ICs
% THD (f_O=1.84 MHz)	0.78-0.84	0.784-0.845	0.79-0.88
Phase Error (f_O=1.84 MHz)	0.21^0	0.21^0	0.24^0
Voltage Supply (V)	±1	±5	±5
Maximum operating frequency (Hz) (shown)	30.58 M	1.94 M	1.84 M
Operating frequency range (Hz) through electronic tuneability; (not shown)	8.14 M-52.94 M	NA	NA
Operating frequency range (Hz) by varying C_2, C_3 (1.8 nF-1.3pF)	822k-52.94M	820k - 1.94 M	820k-1.84M
Frequency sensitivity against tolerances in passive components ($C_{2,3}$)	<0.4%	<0.4%	<0.45%
Power consumption	1.08mW	0.38W	0.4W

VII. Comparison With The Available Literature

The comparison of the reported literature is given in Table V. Designs in [7] are second-order QO realization and use excessive active and passive components. [7] also lacks the electronically tunable feature and % THD is higher. Moreover, the circuit does not provide arbitrary phase and AM signals. Designs in [8] are third-order arbitrary phase shift oscillators (APSOs); however, they employ excessive active blocks and do not work as AM generators. The topologies [10-38] realize TOQO. The topologies [10-30, 32, 37, 38] are only QOs, and [31, 33, 34-36] provide AM along with QO. It is observed that [10-21] use three or more analog building blocks (ABBs), whereas topologies [22-36] use two ABBs. The topologies [37, 38] use only one ABB. Topologies [10-12, 13, 14, 18-19, 22, 23, 26-28, 29, 30] generate either voltage (V) or current (I) signal only. Furthermore, independent electronic control of FO and CO adds an advantage to any circuit realization, but [10, 12-15, 16, 17, 22-24, 29, 31, 32, 38] do not have that property.

The proposed TOQO is completely electronically tunable, for the availability of quadrature current signals, quadrature voltage signals, arbitrary phase-shifted signals (APSO), and AM signals from the same circuit topology. The FO and CO are electronically tuneable and can be controlled independently without affecting each other. Circuit provides amplitude controllability of resulting AM/ASK signals through electronic tuning without affecting the CO and FO of the oscillator. Moreover, a sinusoidal signal having arbitrary phase variation in the range of ± (0 to 90^0) with respect to quadrature current signals is obtained from the proposed configuration. This feature is not available in any of the reported literature. Moreover, the simulated operating frequency range of oscillation of the proposed TOQO is considerably high (820 kHz to 52.94 MHz) compared to other available designs. Furthermore, the phase noise in generated signals in operating range of TOQO is considerably very low (-180 dBc to -240 dBc) as compared to other available values. Further, the phase

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

error between generated quadrature signal is minimal compared to other designs. % THD of voltage and current signals is low (0.84%). Except in [19], layout has not been reported in any designs. It is observed that layout chip area of proposed circuit is very low (0.00781 mm^2) as compared to [19], which occupies total area of 1.44mm^2. Moreover, the power consumption of proposed circuit is also minimum as compared to other designs except in two cases [31, 34] as can be seen from the Table. However, [31] does not provide the independent controlling of CO and FO through electronic tuning. Moreover, both [33] and [34] have low operating frequency and high %THD or high phase error.

TABLE V
COMPARISON WITH THE PREVIOUSLY REPORTED QUADRATURE OSCILLATORS

Ref.	No. and type of Active Block /ABB	No. of Passive element (R/C)	Signal generation types from the circuit	Simultaneous availability of quadrature voltage and current signals	Independent control of FO and CO through electronic tuning	Offer electronic tuneability	Frequency of oscillation (Hz)	THD (%)	Phase error (%,)	Phase noise (dBc)	Power cons-umtion (W)	Lay-out Area (mm^2)
[7]	2 ECCIIs, 1 DVB, 1VA	2/2	QO	No	Yes	No	4 M	1-3	NA	NA	NA	NA
[8]	5 OTAs	0/3	APSO	No	Yes	Yes	1 M	0.6-1.4	NA	NA	NA	NA
[10]	4 DVCC	3/3	TOQO	No	No	Yes	2.5M	0.8-1.8	NA	NA	NA	NA
[11]	4 CCCII	0/3	TOQO	No	Yes	Yes	1.14M	1	NA	NA	NA	NA
[12]	3 Op-Amp	5/3	TOQO	No	No	Yes	Few kHz	NA	NA	NA	NA	NA
		3/5	TOQO	No	No	Yes	NA	NA	NA	NA	NA	NA
[13]	3 OTRA	5/3	TOQO	No	No	Yes	161k	0.57-7	NA	NA	NA	NA
[14]	3 CCII	5/3	TOQO	No	No	Yes	370.3k	NA	NA	NA	NA	NA
		3/5	TOQO	No	No	Yes	215.5k	NA	NA	NA	NA	NA
[15]	3 DVCC	3/3	TOQO	Yes	No	No	7.94M	0.5-1.8	NA	NA	NA	NA
[16]	3 DVCC	3/3	TOQO	Yes	No	Yes	795k	1	1.1	NA	NA	NA
	3VCDVCC	0/3	TOQO	Yes	Yes	Yes	NA	NA	NA	NA	NA	NA
[17]	2CCII, 1 UVC	3/3	TOQO	Yes	No	Yes	1.437M	0.19-0.86	NA	NA	NA	NA
[18]	3 CDTA	0/3	TOQO	No	Yes	Yes	kHz	NA	1.03-2.57	NA	3.948m	NA
[19]	2 CDTA	0/3	TOQO	No	Yes	Yes	73.6M	NA	NA	NA	NA	1.44
	3 CDTA	0/3	TOQO	No	Yes	Yes	212M	NA	NA	NA	NA	
[20]	2 MO-CCII, 1 CCII	5 MOS-R /3	TOQO	Yes	Yes	Yes	3.24M	2-2.94	NA	NA	NA	NA
[21]	3 MO-CCCII	0/3	TOQO	Yes	Yes	Yes	8M	1-2	NA	NA	NA	NA
[22]	2 OTRA	3/4	TOQO	No	No	No	320k	1.8-2.5	3-4.5	NA	NA	NA
[23]	2 OTRA	3/3	TOQO	No	No	Yes	29.04k	1.17	NA	NA	1.52m	NA
		3/3	TOQO	No	No	Yes	5.1k	1.32	NA	NA	1.52m	NA
[24]	2 DV- CCTA	1/3, 2/3	TOQO	Yes	No	Yes	1.59M	2	1.5	NA	3.59	NA
[25]	2 VDCC	3/3	TOQO	Yes	No	Yes	1.12M	<1	NA	NA	1.35m	NA
[26]	2 VDBA	1/3	TOQO	No	Yes	Yes	81.05k	0.4-2.5	NA	NA	NA	NA
[27]	2 VDIBA	1/3	TOQO	No	Yes	Yes	kHz	NA	NA	NA	NA	NA
[28]	1 DDCC, 1 VDTA	1/3	TOQO	No	Yes	Yes	1M	2.95	NA	NA	1.76m	NA
[29]	2 VDCC	3/3	TOQO	Yes	No	Yes	20.87M	1.89	0.37-0.38	-50 to -108	1.28m	NA
[30]	2 VDTA	0/3	TOQO	No	Yes	Yes	1.054M	0.9-1.8	0.9	NA	NA	NA
[31]	2 DVCCTA	2/3	TOQO, AM	Yes	No	Yes	2.11M	< 1.3	< 0.6	-30 to -110	1.028m	NA
[32]	2 FDCCII	4 MOS-R/3	TOQO	Yes	No	Yes	1.56M	< 3	NA	NA	NA	NA
[33]	2 VDTA	0/3	TOQO, AM	No	Yes	Yes	2M	1.36-2.98	NA	NA	NA	NA
[34]	2 VDTA	0/3	TOQO, AM	Yes	Yes	Yes	730k	3.11-3.28	NA	-20 to -105	0.457m	NA

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

[35]	2 MO-CCCTA	0/3	TOQO, AM	Yes	Yes	Yes	2.11M	0.64-1.74	NA	-40 to -100	NA	NA
[36]	2 VDTA	0/3	TOQO, AM	Yes	Yes	Yes	3.17M	0.5-4.5	< 3	-20 to -100	NA	NA
[37]	1 MCCCFTA	0/3	TOQO	Yes	Yes	Yes	2.7M	0.3-0.8	0.3-5.7	NA	1.43m	NA
[38]	1 FDCCII	3/3	TOQO	Yes	No	Yes	0.886M	0.6-1.6	NA	NA	NA	NA
This Work Monolithic CMOS Sim.	1 MDX-CCCTA	0/3	TOQO, AM, APSO ± (0- 90^0)	Yes	Yes	Yes	820k-52.94M	0.78-0.84	< 0.22	-180 to -240	1.08m	0.00781
IC Sim. (Macro Model)	8 AD844s, 3 CA308s	7/3	AT	Yes	NA	No	820k-1.94MHz	0.784-0.845	< 0.22	-182 to -242	0.38	NA
Discrete prototype (Expt.)	8 AD844s, 3 CA308s	7/3	AT	Yes	NA	No	820k-1.84MHz	0.79-0.88	< 0.25	-178 to -238	0.4	NA

*Sim.: Simulation; Expt.: Experimental; AM : Amplitude Modulator; APSO : Arbitrary Phase Shift Oscillator; AT: All three i.e. TOQO, AM, APSO

VIII. Conclusion

This work presents the realization of an electronically tunable TOQO employing a new CM ABB, MDXCCCTA and three grounded passive capacitors. It simultaneously provides two quadrature voltage outputs, two high-impedance quadrature current outputs, and another high-impedance current output with controllable amplitude resulting in AM/ASK modulations. The CO and FO can be controlled independently through electronic tuning. It also provides an arbitrary phase-shifted signal in ± (0 to 90^0) range. The proposed circuit, providing quadrature oscillations, modulation, and phase-shifted signals from the same topology, can be used as a compact module in instrumentation, measurement, and communication techniques. Further, a new proposed automatic stabilization method is implemented in TOQO to control the CO and FO simultaneously. The workability of the proposed TOQO is validated through simulations and experimentation. The oscillator exhibits low %THD (0.78-0.88), low phase error (0.22^0), low power consumption (1.08 mW), low phase noise (-180 dBc to -240 dBc), low sensitivity (<0.45%) against tolerances in passive components and, small deviation (<0.8%) against various PVT corners. Moreover, the proposed circuit possesses a wide electronic and non-electronic tunable range of oscillation frequency from 820 kHz to 52.94 MHz. Furthermore, the proposed TOQO occupies fairly a small layout chip area.

References

[1] S. U. Natarajan, "Measurement of capacitances and their loss factors," *IEEE Transactions on Instrumentation and Measurement*, vol. 38, no. 6, pp. 1083-7, Dec. 1989.

[2] D. R. Bhaskar, and R. Senani, "New CFOA-based single-element-controlled sinusoidal oscillators," *IEEE transactions on instrumentation and measurement*, vol. 55, no. 6, pp. 2014-21, Nov. 2006.

[3] B. Linares-Barranco, T. Serrano-Gotarredona, J. Ramos-Martos, J. Ceballos-Cáceres, J. M. Mora, and A. Linares-Barranco, "A precise 90/spl deg/quadrature OTA-C oscillator tunable in the 50-130-MHz range," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 4, pp. 649-63, Apr. 2004.

[4] H. Hua, Y. Wang, and D. Yan, "A low-cost dynamic range-finding device based on amplitude-modulated continuous ultrasonic wave," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 2, pp. 362-7, Aug. 2002.

[5] T. Addabbo, A. Fort, M. Mugnaini, V. Vignoli, A. Baldi, and M. Bruzzi, "Quartz-crystal microbalance gas sensors based on TiO 2 nanoparticles," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 3, pp. 722-30, Jan. 2018.

[6] F. Aydemir and M. A. Ebeoğlu, "A QCM sensor array-based electronic tongue with the optimized oscillator circuit using FPGA," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 2, pp. 431-8, Dec. 2017.

[7] R. Sotner, J. Jerabek, N. Herencsar, and J. Petrzela, "Methods for extended tunability in quadrature oscillators based on enhanced electronic control of time constants," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 6, pp. 1495-505, 2018.

[8] S. H. Tu, Y. S. Hwang, J. J. Chen, A. M. Soliman, and C. M. Chang, "OTA-C arbitrary-phase-shift oscillators," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2305-19, Feb 2012.

[9] B. Wilson, "Recent developments in current conveyors and current-mode circuits," *IEEE Proceedings (Circuits, Devices and Systems)*, vol. 137, no. 2, pp. 63-77, Apr. 1990.

[10] S. Maheshwari, "Analogue signal processing applications using a new circuit topology," *IET Circuits, Devices & Systems*, vol. 3, no. 3, pp. 106-15, Jun. 2009.

[11] S. Maheshwari and I. A. Khan, "Current controlled third-order quadrature oscillator," *IEEE Proceedings-Circuits, Devices and Systems*, vol. 152, no. 6, pp. 605-7, Dec. 2005.

[12] J. W. Horng "Quadrature oscillators using operational amplifiers," *Active and Passive Electronic Components*, Jan. 2011; Article ID 320367.

[13] R. Pandey, N. Pandey, G. Komanapalli, and R. Anurag, "OTRA based voltage-mode third-order quadrature oscillator," *ISRN Electronics*, Apr. 2014; Article ID 126471.

[14] J. W. Horng, C. L. Hou, C. M. Chang, W. Y. Chung, H. W. Tang, and Y. H. Wen, "Quadrature oscillators using CCIIs," *International Journal of Electronics*, vol. 92, no. 1, pp. 21-31, Jan. 2005.

[15] B. Chaturvedi and S. Maheshwari, "Third-order quadrature oscillator circuit with current and voltage outputs," *ISRN Electronics* Aug. 2013; Article ID 385062.

[16] S. Maheshwari, "Quadrature oscillator using grounded components with current and voltage outputs," *IET Circuits, Devices & Systems*, vol. 3, no. 4, pp. 153-60, Aug. 2009.

[17] J. Koton, N. Herencsar, K. Vrba, and B. Metin, "Current-and voltage-mode third-order quadrature oscillator," In *13th International IEEE Conference on Optimization of Electrical and Electronic Equipment (OPTIM)* May. 2012, pp. 1203-1206.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3300449

[18] J. W. Horng, "Current-mode third-order quadrature oscillator using CDTAs, "*Active and Passive Electronic Components,"* Jan. 2009; Article ID 789171.

[19] J. Jin, C. Wang, and J. Sun, "Novel third-order quadrature oscillators with grounded capacitors," *Automatik*a, vol. 56, no. 2, pp. 207-16, Jan. 2015.

[20] H. Cicekli and A. Gokcen "MOS-C Based Electronically Tuneable Current/Voltage-Mode Third Order Quadrature Oscillator and Biquadratic Filter Realization," *Elektronika ir Elektrotechnika*, vol. 27, no. 3, pp. 38-49, Jun. 2021.

[21] S. Maheshwari, "Current-mode third-order quadrature oscillator," *IET Circuits, Devices & Systems,* vol. 4, no. 3, pp. 188-95, May 2010.

[22] G. Komanaplli, N. Pandey, and R. Pandey, "New realization of quadrature oscillator using OTRA," *International Journal of Electrical and Computer Engineering,* vol. 7, no. 4, pp. 1815, Aug. 2017.

[23] B. C. Nagar and S. K. Paul, "Voltage mode third-order quadrature oscillators using OTRAs," *Analog Integrated Circuits and Signal Processing,* vol. 88, no. 3, pp. 517-30, Sep. 2016.

[24] N. Pandey and R. Pandey, "Approach for third-order quadrature oscillator realization," *IET Circuits, Devices & Systems,* vol. 9, no. 3, pp. 161-71, May 2015.

[25] S. Roy, T. K. Paul, and R. R. Pal, "Realization of third-order voltage-mode/current-mode quadrature oscillator circuit employing VDCCs and all grounded capacitors," *Journal of Physical Sciences,* vol. 23, pp. 131-145, Dec. 2018.

[26] K. L. Pushkar, "Voltage-mode third-order quadrature sinusoidal oscillator using VDBAs," *Circuits and Systems,* vol. 8, no. 12, pp. 285-92, Dec. 2017.

[27] K. L. Pushkar and D. R. Bhaskar, "Voltage-mode third-order quadrature sinusoidal oscillator using VDIBAs," *Analog Integrated Circuits and Signal Processing,* vol. 98, no. 1, pp. 201-7, Jan. 2019.

[28] K. Phanruttanachai and W. Jaikla, "Third-order current-mode quadrature sinusoidal oscillator with high output impedances," *International Journal of Electronics and Communication Engineering,* vol. 7, no. 3, pp. 300-3, Mar. 2013.

[29] S. Roy and R. Pal, "Electronically tunable third-order dual-mode quadrature sinusoidal oscillators employing VDCCs and all grounded components," *Integration,* vol. 76, pp. 99-112, Jan. 2021.

[30] P. Phatsornsiri, P. Lamun, M. Kumngern, and U. Torteanchai, "Current-mode third-order quadrature oscillator using VDTAs and grounded capacitors," In *4th Joint International IEEE Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE),* Mar. 2014, pp. 1-4.

[31] H. P. Chen, Y. S. Hwang, and Y. T. Ku, "A systematic realization of third-order quadrature oscillator with controllable amplitude," *AEU-International Journal of Electronics and Communications,* vol. 79, pp. 64-73, Sep. 2017.

[32] J. Mohan, B. Chaturvedi, and A. Kumar, "Active-C Realization of multifunction biquadratic filter and third-order oscillator," *Radio Science,* vo. 55, no. 1, e2019RS006877, Jan. 2020.

[33] O. Channumsin and A. Jantakun, "Third-order sinusoidal oscillator using VDTAs and grounded capacitors with amplitude controllability," In *4th Joint International IEEE Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE),* Mar. 2014, pp. 1-4.

[34] H. P. Chen, S. F. Wang, Y. N. Chen, and Q. G. Huang, "Electronically tunable third-order quadrature oscillator using VDTAs," *Journal of Circuits, Systems and Computers,* vol.28, no.4, pp. 1950066, Apr. 2019.

[35] H. P. Chen, Y. S. Hwang, and Y. T. Ku, "Voltage-mode and current-mode resistorless third-order quadrature oscillator," *Applied Sciences,* vol. 7, no. 2, pp. 179, Feb. 2017.

[36] H. P. Chen, Y. S. Hwang, and Y. T. Ku, "A new resistorless and electronic tunable third-order quadrature oscillator with current and voltage outputs," *IETE Technical Review,* vol. 35, no. 4, pp. 426-38, Jul. 2018.

[37] K. Khaw-Ngam, M. Kumngern, and F. Khateb, "Mixed-mode third-order quadrature oscillator based on single MCCFTA," *Radioengineering,* vol. 26, no. 2, pp. 522-35, Jun. 2017.

[38] E. Wareechol, B. Knobnob, and M. Kumngern, "FDCCII-Based Third-Order Quadrature Sinusoidal Oscillator," In *41st International IEEE Conference on Telecommunications and Signal Processing (TSP),* Jul. 2018, pp. 1-4.

[39] A. Zeki and A. Toker, "The dual-X current conveyor (DXCCII): a new active device for tunable continuous-time filters," *International Journal of Electronics,* vo. 89, no. 12, pp. 913-23, Dec. 2003.

[40] S. A. Tekin, H. Ercan, and M. Alçi, "A versatile active block: DXCCCII and tunable applications," *Radioengineering,* vol. 23, no. 4, pp. 1130-9, Dec. 2014.

[41] S. Maheshwari and J. Rajpoot, "Electronically tunable Quadrature Oscillator circuit using DXCCCII," In *International IEEE Conference on Electrical, Electronics and Computer Engineering (UPCON),* Nov. 2019, pp. 1-5.

[42] B. Chaturvedi, A. Kumar, and J. Mohan, "Low voltage-operated current-mode first-order universal filter and sinusoidal oscillator suitable for signal processing applications," *AEU-International Journal of Electronics and Communications,* vo. 99, pp. 110-8, Nov. 2019.

[43] B. Chaturvedi and A. Kumar, "Fully electronically tunable and easily cascadable square/triangular wave generator with duty cycle adjustment," *Journal of Circuits, Systems and Computers,* vol. 28, no. 4, pp. 1950105, Jun. 2019.

[44] N. Kumar, J. Vista, and A. Ranjan, "A tunable active inductor employing DXCCTA: grounded and floating operation," *Microelectronics Journal*, vol. 90, pp. 1-1, Aug. 2019.

[45] L. Xuechu, G. Qingyun, Q. Shicai, "A low-power high-frequency CMOS peak detector," *Journal of Semiconductors,* vol. 27, no. 10, pp. 1707-10, Oct. 2006.

[46] A. Raj, D. R. Bhaskar, and P. Kumar, "Two quadrant analog voltage divider and square-root circuits using OTA and MOSFETs," *Circuits, Systems, and Signal Processing*, vol. 39, no. 12, pp. 6358-85, Dec. 2020.

[47] S. I. Liu, "Square-rooting and vector summation circuits using current conveyors," *IEE Proceedings-Circuits, Devices and Systems,* vol. 142, no. 4, pp. 223-6, Aug. 1995.

[48] S. Menekay, R. C. Tarcan, and H. Kuntman, "Novel high-precision current-mode multiplier/divider," *Analog Integrated Circuits and Signal Processing,* vol. 60, no. 3, pp. 237-48, Sep. 2009.

Garima Shukla did her B. Tech. in Electronics and Communication Engineering from S. V. C. E. Indore. She received M. Tech. degree in Embedded Systems and VLSI Design from SVVV, Indore and did her M. Tech. project from Indian Institute of Technology (IIT), Indore. She is currently a Ph. D. research scholar in the Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad. Her area of research includes analog VLSI circuit design using current mode approach and CMOS analog signal processing.

Sajal K. Paul did his B.Tech., M.Tech., and Ph.D. in Radio Physics and Electronics from the Institute of Radio Physics and Electronics, University of Calcutta. He has served Webel Telecommunication Industries, Kolkata; Indira Gandhi National Open University (IGNOU), Kolkata; Advanced Training Institute for Electronics & Process Instrumentation (ATI-EPI), Hyderabad; North Eastern Regional Institute of Science & Technology (NERIST), Nirjuli and Delhi College of Engineering (DCE), Delhi in various capacities. He has served the Department of Electronics Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, as Head of the Department, Dean (Faculty), and at present, Professor of the same institute. His research interest includes Microelectronic Devices, Electronic properties of Semiconductor, and Bipolar and MOS Analog Integrated Circuits. Dr. Paul has more than 142 publications in International and National Journals of repute and conferences.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:42:20 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

2023 IEEE 3rd International Conference on Software Engineering and Artificial Intelligence (SEAI) | 979-8-3503-3795-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/SEAI59139.2023.10217367

Towards Accurate Stress Classification: Combining Advanced Feature Selection and Deep Learning

Lovish Arya
Department of Electronics and Communication Engineering
Delhi Technological University
New Delhi, India
lovisharya_2k19ec102@dtu.ac.in

Harshal Chowdhary
Department of Electronics and Communication Engineering
Delhi Technological University
New Delhi, India
harshalchowdhary_2k19ec071@dtu.ac.in

Ishita Agrawal
Department of Electronics and Communication Engineering
Delhi Technological University
New Delhi, India
ishitaagrawal_2k19ec077@dtu.ac.in

Indu Sreedevi
Department of Electronics and Communication Engineering
Delhi Technological University
New Delhi, India
s.indu@dtu.ac.in

***Abstract*— Stress is a significant health issue that affect both physical and mental health. The application of deep learning and machine learning techniques for physiological signal processing has become popular in detecting and classifying stress. In this research, we developed a fresh approach to classify stress on multiple levels using advanced machine and deep learning models. The PhysioNet SRA database was used to collect physiological signals which were pre-processed using a pipeline of filters and normalisation techniques to remove noise and artefacts. These signals were subjected to various feature extraction algorithms to determine the most pertinent characteristics for categorisation. On the acquired features, dimensionality reduction techniques were applied in order to further improve classification performance. The evaluation of Random Forest model outscored all the other classic ML models, achieving an accuracy of 93.77% and 82.42% in classifying stress on two levels and three levels respectively. Among the DL models, the LSTM model is most effective, achieving an accuracy of 95.67% and 88.70% in classifying stress on two levels and three levels respectively. The proposed approach provides a more comprehensive analysis of stress levels, which can help individuals manage stress more effectively.**

***Keywords*—Classification, Deep learning, Machine learning, Physiological signals, Stress detection.**

I. INTRODUCTION

Stress has become a pervasive problem in modern so- ciety, with the potential to have significant adverse effects on both physical and mental health. The recognition and early detection of stress are crucial in preventing its negative consequences. In order to identify and categorise stress based on physiological signals, there has been an increase in interest in using deep learning and machine learning approaches. In this work, we provide a brand-new model for categorising stress on different scales utilising advanced deep and machine learning techniques, resulting in a more thorough examination of stress levels.

The traditional methods for detecting and classifying stress are based on self-reporting by individuals, which can be subjective and prone to errors [1]. However, A more objec- tive and accurate method for stress detection is provided by physiological signals like the electrocardiogram (ECG), pho- toplethysmography (PPG), and galvanic skin response (GSR). The application of deep learning and machine learning models for the classification of stress on multiple levels provides a more nuanced analysis of stress levels that can help individuals manage their stress more effectively.

Previous studies have primarily focused on binary classifi- cation of stress, with only a few attempts to classify stress on multiple levels. The proposed approach in this study provides a more comprehensive analysis of stress levels, which is crucial in providing a more detailed analysis of stress. Stress is a complex phenomenon that involves different levels of severity, and the traditional binary classification approach does not provide a complete understanding of stress levels.

To achieve the classification of stress on multiple levels, we utilised the PhysioNet SRA database, which contains phys- iological signals collected from individuals in various stress conditions [2]. The database provides a wide range of stress levels, making it an appropriate choice for the classification task. The signals were pre-processed using a pipeline of filters and normalisation techniques to remove noise and artefacts [2]. A bandpass filter was used to remove high and low-frequency noise, while a notch filter was used to remove power line interference. After that, the signals were standardised using Z-score standardization to remove the impact of various scales on the categorisation [2].

To find the characteristics that would be most useful for classification, the pre-processed signals were exposed to a variety of feature extraction approaches. Techniques for feature

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:00:42 UTC from IEEE Xplore. Restrictions apply.

selection were also used to decrease the dimensionality of the extracted features, further improving the classification performance [3]. The study evaluated several Machine learning models, including random forest (RF), logistic regression (LR), k-nearest neighbours (KNN), decision tree, and support vector machine (SVM). The deep learning models evaluated in this study include long short-term memory (LSTM) networks, convolutional neural networks (CNN) [4], and hybrid models that combine CNNs and LSTMs.

This study's innovation is based on the use of advanced deep learning models to classify stress on multiple levels. The proposed approach in this study provides a more comprehensive analysis of stress levels and can help individuals manage their stress more effectively. The study's findings demonstrate that the RF model outperforms the competition by reaching an accuracy of 93.77% in classifying stress on 2 levels and 82.42% in classifying stress on 3 levels. Among the deep learning models, the LSTM model performs the best, achieving an accuracy of 95.67% in classifying stress on 2 levels and 88.7% in classifying stress on 3 levels. These results demonstrate the potential of advanced deep learning and machine learning models in the classification of stress on multiple levels.

The study's suggested methodology offers a fresh and useful way to categorise stress on various levels. The use of advanced machine and deep learning models provides a more comprehensive analysis of stress levels, which can help individuals manage their stress more effectively. The findings of this research lay the groundwork for further research in the subject and show the possibility of the suggested methodology to increase the accuracy of stress classification.

II. Methodology

The approach for the features for stress classification is covered in this section.

A. Database Information

In this work, stress level detection was carried out using the SRA Database, which Healey and Picard established [2]. It is a freely available dataset that contains the .dat and .hea files of multiparameter recordings made by 17 healthy Boston, USA, drivers [2]. Fig. 1 illustrates the setup used for the collection of data. It covers signals obtained from drivers who experienced varying degrees of stress brought on by driving on highways, city streets with traffic signals, and unanticipated risks brought on by cyclists [2]. Each .dat file in the dataset contains four different type of physiological signals, including EMG (Electromyography), GSR (Galvanic Skin Response), ECG (Electrocardiogram), and Respiration. EMG and ECG have sample rates of 15.5 Hz and 496 Hz, respectively, but hand, foot, and RESP signals all have sampling rates of 31 Hz [2]. The dataset is an essential resource for research in stress level detection, providing high-quality recordings that enable the development of accurate and reliable stress level detection algorithms [2].

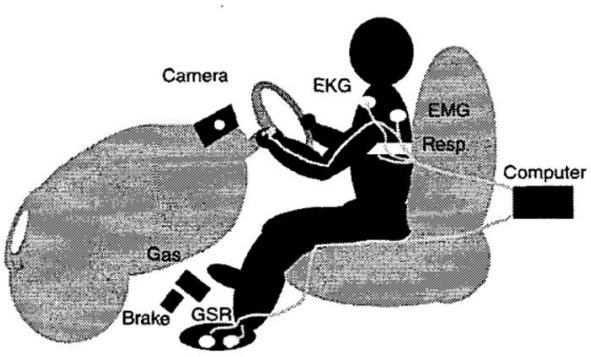


Fig. 1. SRA Data Collection Technique [6]

Fig. 2. Pan-Tompkins Algorithm Methodology

B. Pre-processing

ECG signals are prone to noise and artefacts, therefore, we used various pre-processing techniques to ensure high-quality signal input to our algorithms. Our pre-processing pipeline consisted of the following steps:

1) Extracting RR Intervals from ECG: We initially identified the R-R intervals from each ECG signal. In order to do this, we used a QRS complex detection algorithm based on template matching. We utilized the QRS complex locations to detect the R peaks and calculated the time intervals between them to obtain the RR intervals.

2) Peak detection: We used a method that S. Chaudari et al. had previously used [5]. Peak detection by template matching is a crucial step in detecting R peaks in the ECG signals. One of the widely used algorithms for detecting the R peak is the Pan-Tompkins algorithm. However, this algorithm may not work efficiently for all ECG signals, particularly in the presence of noise and baseline wander. Therefore, to enhance the R peak detection process, we utilized template matching in our study. Fig 2 illustrates the pre-processing phase of the Pan–tompkins algorithm using a block diagram.

Template matching is a technique used to compare a given signal segment with a known template signal to identify similarities. In our study, we used a reference template signal for the R peak in the ECG signal. This reference template was created by averaging the R peak segments from multiple ECG signals to obtain a representative R peak signal. It involves moving the reference template signal along the ECG signal, and at each position, the similarity between the ECG signal segment and the reference template signal was calculated using cross-correlation. In the ECG signal, the R peak was located where the cross-correlation value's peak was at.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:00:42 UTC from IEEE Xplore. Restrictions apply.

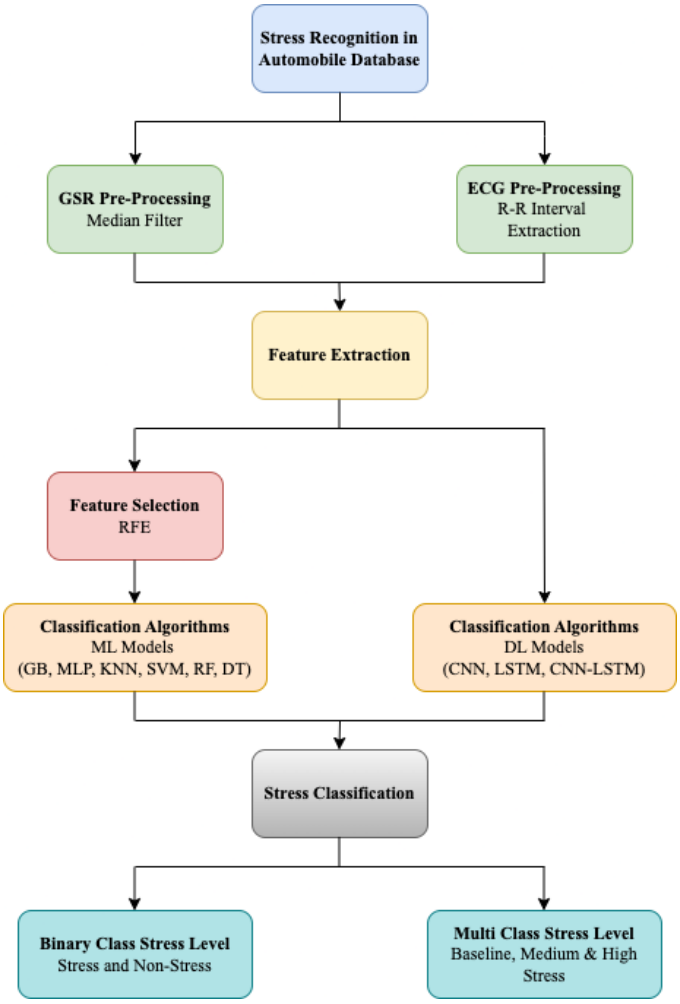


Fig. 3. Stress Classification Methodology

As per V Nivitha Varghees et al [7], to reduce the effects of noise and baseline wander on the cross-correlation, we also applied a pre-processing step that involved baseline correc- tion and band-pass filtering to the ECG signals. After pre-processing, we applied the template matching technique to detect the R peaks in the ECG signal. The use of template matching in conjunction with the Pan-Tompkins algorithm significantly improved the R peak detection accuracy, particularly in the presence of noise and baseline wander. This approach allowed for better identification of the R peaks and helped to reduce the false detection of R peaks. The accurate identification of the R peak is crucial in the computation of the RR interval, which is an essential feature in the detection of stress levels in our study.

3) RR interval Correction: After detecting R peaks in the ECG signals, the next step is to calculate the RR intervals between consecutive R peaks. However, in some cases, the detected RR intervals may not be accurate due to various factors such as noise, artifacts, and outliers. To address this issue, a correction method is used to ensure that the RR intervals are accurate and reliable.

In this paper, the RR interval correction was performed using a median filter. The median filter is a non-linear digital signal processing technique that replaces each sample value with the median value of neighboring samples. This technique is widely used to remove outliers from signals and to smooth noisy signals. To apply the median filter for RR interval correction, we first calculated the initial RR intervals from the detected R peaks. The Z-score, which is the difference between each RR interval and the distribution's mean divided by the standard deviation, was then determined for each RR interval [8]. Any RR interval with a Z-score greater than 2 was considered an outlier and replaced with the median value of the neighboring RR intervals.

After applying the median filter, the corrected RR intervals were obtained, which were then used to extract features for stress level classification. This correction method helped to remove any outliers and ensure that the RR intervals were accurate and reliable, thus improving the overall performance of the stress level detection system.

4) Stress Labeling And Windowing: The ECG signals were labeled based on the driver's stress level during the recording. This was done by assigning a stress level of 1 for a relaxed state and a stress level of 2 for a stressed state. The stress level was determined based on the driver's self-reported stress level during the recording. After labeling the ECG signals, the next step was to divide them into 5-minute windows. Each 5-minute window was assigned a stress level based on the majority stress level of the ECG beats in the window. This ensured that each window was accurately labeled and helped in creating a balanced dataset for training the model.

The suggested model was made to recognise three distinct drive phases: Low, Middle, and High Stress. We utilized the marker signals from the dataset to achieve distinctive. The low-stress data was considered to be between 700 to 1000 seconds, medium-stress data between 1600 to 1900 seconds, and high-stress data between 3200 to 3500 seconds [2]. Once the data was labeled and the stress levels were identified, the next step was windowing. The model separated the signal into
0.5 min frames with 0.25 min overlaps every 5 min for each stress level. For each stress label, this led to 19 splits from each participant. For instance, we obtained 980 splits altogether for multi-level stress (baseline, mid, and high-level) and about 570 splits for binary stress (stress and non-stress) [2].

C. Feature Extraction

Various characteristics were taken from the ECG data to indicate the physiological condition of the driver [9]. Time-domain, frequency-domain, and nonlinear characteristics were among these features. Included in the list of time-domain characteristics were the RR intervals' mean, standard deviation, and coefficient of variation. According to A. D. Wibawa et al. [10], The frequency-domain features included power spectral density in the low-frequency (LF, 0.04-0.15 Hz) and high- frequency (HF, 0.15-0.4 Hz) bands, as well as the LF/HF ratio. The nonlinear features included sample entropy, approximate entropy, and detrended fluctuation analysis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:00:42 UTC from IEEE Xplore. Restrictions apply.

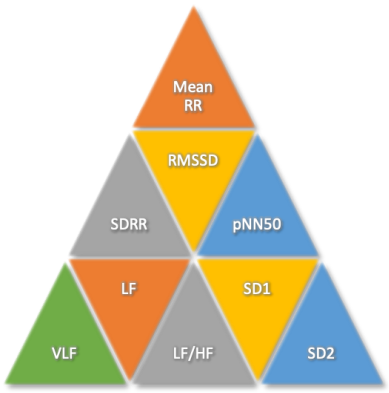


Fig. 4. Advance features selected after RFE

D. Feature Selection

Recursive Feature Elimination (RFE), a sophisticated feature selection method, was used in this work. RFE is a feature selection technique based on wrappers that eliminates features repeatedly while building a model on the remaining features. The method ranks the features according to their importance and eliminates the least important feature at each iteration until the desired number of features is reached. The full feature set [11] was used to train a support vector machine (SVM) classifier with a radial basis function (RBF) kernel for RFE [11]. The feature significance scores were then calculated by the classifier using the coefficients or weights assigned to the features [11]. The model was retrained using the smaller feature set after removal of the least significant feature until the necessary quantity of characteristics was reached. A cross-validation approach was used to set the number of features to be selected at 9.

The top 9 features selected through RFE were Mean RR, SDRR, RMSSD, pNN50, VLF, LF, LF/HF, SD1, and SD2. Fig. 4 demonstrates the 9 features that were selected. These features were chosen based on their ability to distinguish strained from non-stressed conditions, as well as their relevance to the physiological mechanisms underlying stress [12]. These include non-linear, time domain, and frequency dcomain properties, all of which have been demonstrated to be useful in earlier research on stress detection. The use of RFE allowed for the identification of a set of informative features that could be used in the development of more effective stress detection algorithms and interventions for individuals at risk of stress related disorders.

E. Parameter Tuning

In order to fine-tune the classifier variables, we employed a grid search strategy. We tested the performance of the classifiers with several parameter combinations and selected the parameter combination that generated the best result. We changed the parameters for the support vector machine (SVM), k-nearest neighbour (KNN), and random forest (RF) classifiers.

F. Classification Algorithms

This study used biosignals from ECG-based heart rate variability (HRV) and galvanic skin reaction (GSR) to classify stress using various machine learning (ML) and deep learning (DL) methods. The objective was to enhance the accuracy of stress level classification by exploring advanced algorithms.

1) Machine Learning: Seven machine learning (ML) models were tested, including Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF), k-Nearest Neighbor (k-NN), Logistic Regression (LR), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP).

Following hyperparameter tuning and K-fold Cross-Validation, we discovered that the RF model had the best performance, with an 93.77 accuracy%, an 0.89 F1-score, and an 0.94 area under the curve (AUC). The other models had lower accuracy scores, ranging from 76% for LR to 93% for RF.

2) Deep Learning: In addition to ML models, we also examined the use of DL algorithms for stress classification. We tested three different DL models: Convolutional Neural Network (CNN) [4], Long Short-Term Memory (LSTM), and a combination of CNN-LSTM.

Our results indicated that DL models outperformed ML models, with the LSTM model achieving the best performance. The LSTM model has an 0.96 F1-score, an 0.96 AUC, and an 95.67%. accuracy. outperforming both the RF model and all other ML models. The CNN [4] model and the combination of CNN-LSTM had lower accuracy scores of 94.2% and 94.75%, respectively.

Overall, our findings suggest that DL algorithms, especially the LSTM model, are effective for stress classification using HRV and GSR biosignals. The higher accuracy scores obtained by DL models compared to ML models highlight the potential for DL algorithms to improve stress detection and management technologies.

III. Results and Discussions

This work used HRV and GSR bio-signals extracted from the ECG to increase the precision of stress level categorisation using machine and deep learning methods. The individuals who drove in various stressful scenarios provided the data.

The parameters of the various classifier models were optimised via hyperparameter tuning. K-fold Cross-Validation was performed to validate the performance of the models in order to guarantee objective outcomes. Out of 32 features, 9 features were selected based on their high correlation with stress level changes. The results showed that HRV and GSR signals are reliable indicators of stress level changes. These results are in line with other research that showed how stress affects health on HRV and GSR signals. However, there may be limitations to their study, such as the sample size and the stress situations tested.

In the current work, a range of deep learning and machine learning methods were used to precisely identify stress levels using physiological data. The informative features were extracted from the physiological signals via feature extraction

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:00:42 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARATIVE ANALYSIS OF RELATED WORK

Related Work	Signals	Techniques	Window size (in min.)	Accuracy (%)	Stress Levels
Elgendi et al. [13]	HRV, HR, EMG, GSR, RESP	ML Models (KNN, DT, LR, SVM, QDA)	0.25, 0.5	75	Binary
Pankajavalli et al. [14]	HRV, HR, EMG, GSR	ML Models (SVM, NB, DT, LR, ANN, KNN)	0.5	88.75	Binary
Lopez-Martinez et al. [15]	HR and EMG	ML Models (LR, SVM)	0.5	92.85	Binary
Jime´nez-Limas et al. [16]	HR, RESP, HRV, FGSR	ML Models (SVM, KNN, LR, RF, NB)	05	81.4	Binary
Dalmeida et al. [17]	ECG	ML Models (GB, MLP, NB, RF, KNN, SVM)	0.5	76.53	Binary
Sugested Model	**ECG and GSR**	**DL Models (CNN, LSTM, CNN-LSTM), ML Models (GB, DT, MLP, KNN, SVM, RF, LR)**	**0.5**	**95.67**	**Binary and Multi level**

TABLE II
PREDICTION OF ML AND DL TECHNIQUES, CHARACTERISING STRESS IN BINARY AND MULTI-CLASS

Technique	Binary-Level Stress		Multi-Level Stress	
	F1 Score (0-1)	Accuracy	F1 score (0-1)	Accuracy
CNN	0.93	94.2 %	0.87	85.60 %
LSTM	**0.96**	**95.67** %	**0.90**	**88.70** %
CNN-LSTM	0.95	94.75 %	0.88	87.56 %
SVM	0.79	83.92 %	0.68	64.93 %
Gradient Boosting	0.84	88.89 %	0.68	70.03 %
Logistic Regression	0.68	76.32 %	0.50	52.45 %
MLP	0.82	80.05 %	0.54	53.95 %
Random Forest	0.89	93.77 %	0.83	82.42 %
Decision Tree	0.86	91.67 %	0.82	79.52 %
KNN	0.80	87.83 %	0.77	76.12 %

and selection techniques, and the performance of the models was evaluated using various evaluation metrics. The results indicated that in terms of properly identifying stress levels, deep learning models fared better than machine learning models. The LSTM model outperformed all other evaluated models, with an accuracy of 95.67%, an 0.96 F1-score, and an 0.96 AUC.

These results are in line with earlier research that shown the potency of deep learning methods for stress detection. For instance, Chen et al. (2018) also used physiological signals for stress detection and achieved an accuracy of 87.7% using a deep learning model. Likewise, Zhang et al. (2020) employed a combination of physiological signals and facial expressions for stress detection and attained an accuracy of 92.1% using a deep learning model. These studies support the notion that deep learning techniques are effective in accurately detecting stress levels via physiological signals. The implications of our study are significant in the development of stress detection and management technologies, which can help individuals monitor and control their stress levels in real-time. The accurate detection of stress levels via physiological signals enables the timely delivery of stress management interventions and improves overall well-being.

To summarize, our study demonstrated the superiority of deep learning techniques over machine learning techniques in accurately detecting stress levels using physiological signals. The results align with prior research and have impor- tant implications for the development of stress management technologies. Further studies are required to see whether our conclusions can be applied to different populations and environments.

Table I compares the best outcomes from recent work that have been examined. The authors discard intermediate (Medium) stress evaluation in favour of a binary classification (Low-High). Our research, however, focuses on multi-level stress labeling trained on advanced deep learning model.

The outcomes of the suggested work are shown in Table II. Considering F1 scores 0.96 and 0.90 for the multi-level and binary labelling tasks, respectively, suggested through this dataset, these results demonstrate that LSTM outperforms competing models.

$$F1score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \quad (1)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:00:42 UTC from IEEE Xplore. Restrictions apply.

IV. Conclusion

This study aimed to improve stress level classification utilising deep learning and machine learning techniques with HRV and GSR bio-signals. Different ML techniques, including SVM, KNN, LR, Decision Tree, Random Forest and NB, were evaluated along with three different DL models, including CNN, LSTM, and a combination of CNN-LSTM. After applying hyperparameter tuning and K-fold Cross-Validation, we found that DL models outperformed ML models, particularly the LSTM model which achieved an 95.67% accuracy, 0.96 F1-score, and 0.96 AUC, demonstrating the potential for DL algorithms to improve stress detection and management technologies. Our findings suggest that the use of HRV and GSR bio-signals with DL algorithms, particularly LSTM, can be effective in accurately classifying stress levels.

The study provides a promising avenue for further research in this area, particularly in exploring the use of additional bio-signals and expanding the sample size. Future studies should consider incorporating other bio-signals and expanding the sample size to validate our findings further. In addition, it is recommended to compare and benchmark our results with recent studies in the area.

Overall, the study contributes to the growing body of research on stress classification and management, particularly in exploring the potential of DL algorithms to improve stress detection technologies. The study has significant implications for stress management in various domains, including healthcare and workplace environments.

References

[1] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 440–460, 2019.

[2] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.

[3] Y. Wang, X. Jiang, K. Yu, F. Shi, L. Qin, H. Zhou, and F. Cai, "Infrared thermal images classification for pressure injury prevention incorporating the convolutional neural networks," *IEEE Access*, vol. 9, pp. 15 181–15 190, 2021.

[4] D. Dai, "An introduction of cnn: models and training on neural network models," in *2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR)*. IEEE, 2021, pp. 135–138.

[5] S. Chaudhuri, T. D. Pawar, S. Duttagupta, S. Chaudhuri, S. Duttagupta, and T. D. Pawar, "Activity recognition," *Ambulation Analysis in Wearable ECG*, pp. 79–121, 2009.

[6] J. Healey and R. Picard, "Smartcar: detecting driver stress," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 4. IEEE, 2000, pp. 218–221.

[7] V. Nivitha Varghees, H. Cao, and L. Peyrodie, "Variational mode decomposition based simultaneous r peak detection and noise suppression for automatic ecg analysis," *IEEE Sensors Journal*, pp. 1–1, 2023.

[8] M. Amin, K. Ullah, M. Asif, A. Waheed, S. U. Haq, M. Zareei, and R. R. Biswal, "Ecg-based driver's stress detection using deep transfer learning and fuzzy logic approaches," *IEEE Access*, vol. 10, pp. 29 788–29 809, 2022.

[9] P. A. Va´squez-Ucho, R. Valencia-Ramos, F. Villalba-Meneses, A. Tirado-Esp´ın, V. H. Salazar, and D. Almeida-Gala´rraga, "Development of a system to detect stress using electrocardiographic signals and machine learning models," in *2022 Third International Conference on Information Systems and Software Technologies (ICI2ST)*, 2022, pp. 57–63.

[10] M. J. Rodrigues, O. Postolache, and F. Cercas, "The influence of music stimulation on heart rate variability: Preliminary results," in *2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2022, pp. 1–6.

[11] T. Anwar and S. Zakir, "Machine learning based real-time diagnosis of mental stress using photoplethysmography," in *Journal of Biomimetics, Biomaterials and Biomedical Engineering*, vol. 55. Trans Tech Publ, 2022, pp. 154–167.

[12] P. A. Va´squez-Ucho, R. Valencia-Ramos, F. Villalba-Meneses, A. Tirado-Esp´ın, V. H. Salazar, and D. Almeida-Gala´rraga, "Development of a system to detect stress using electrocardiographic signals and machine learning models," in *2022 Third International Conference on Information Systems and Software Technologies (ICI2ST)*, 2022, pp. 57–63.

[13] M. Elgendi *et al.*, "Machine learning ranks ecg as an optimal wearable biosignal for assessing driving stress," *IEEE Access*, vol. 8, pp. 34 362–34 374, 2020.

[14] P. Pankajavalli *et al.*, "A unified framework for stress forecasting using machine learning algorithms," *Advances in Computational Intelligence and Informatics. ICACII*, 2019.

[15] D. Lopez-Martinez *et al.*, "Detection of real-world driving-induced affective state using physiological signals and multi-view multi-task machine learning," in *2019 8th International Conference on Affective Com- puting and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2019, pp. 356–361.

[16] M. A. Jime´nez-Limas *et al.*, "Feature selection for stress level classification into a physiologycal signals set," in *2018 15th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*. IEEE, 2018, pp. 1–5.

[17] K. M. Dalmeida *et al.*, "HRV features as viable physiological markers for stress detection using wearable devices," *Sensors*, vol. 21, no. 8, p. 2873, 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on September 04,2023 at 07:00:42 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023)
IEEE Xplore Part Number: CFP23AWO-ART; ISBN: 979-8-3503-9663-8

2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192687

Tuning A PID Controller using Metaheuristic Algorithms

Mini Sreejeth[a], Rishant Kumar[b], Nishant Tripathi[c], Prateek Garg[d]
Department of Electrical Engineering
Delhi Technological University
Delhi, India

[a]minisreejeth@dtu.ac.in, [b]rishantkumarimp510@gmail.com, [c]ashutripathi12343@gmail.com
[d] prateek.pgarg@gmail.com

***Abstract*—This study describes the difficulties that arise while fine-tuning the PID controllers. PID controllers are used in various systems to keep things stable but finding the right settings can be tricky, especially for complex systems that behave in unpredictable ways. If the controller has not been set up correctly, it can cause the system to become unstable and not perform as intended. Although there are some techniques that have shown promise in the past, they also have their limitations.**

To tackle this issue, this study proposes a solution that uses metaheuristic algorithms such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Ant Colony Optimization (ACO). These algorithms are designed to find optimal settings for the controller more efficiently than traditional methods. The objective of this research is to compare how well these new methods are compared to the traditional methods. The results show that the new algorithms are more effective and lead to better performance and stability in the system. This research provides a promising solution to the challenges of fine-tuning PID controllers, especially for more complex systems that are difficult to control.

Keywords*— *PID Controller, Optimization, Genetic Algorithm, Ant Colony Optimization, Particle Swarm Optimization, Rise Time, Maximum Overshoot, Settling Time

I. INTRODUCTION

PID controller is a crucial type of controller that has found widespread use in various industry segments. In most industrial applications, PID controllers offer excellent static and dynamic responses because of their simplistic structure and strong performance under a broad range of operating situations [1][2].

In the past, engineers typically used a trial-and-error approach to handle the matter. They manually tuned PID controllers, which was a time-consuming process. Various methods can be found, with the ZN tuning formula being the most well-known method. Once tuned using the ZN method, a good system response can be achieved, but not necessarily an optimal one.

On the other hand, Genetic Algorithm (GA) is highly effective at finding solutions to a variety of problems. ACO is a versatile optimization method and PSO's ability to work with swarms allows it to explore multiple local minima, increasing the chances of finding the global minima.

A. BACKGROUND

Gas turbines are commonly used in various industries, such as refineries and chemical plants, to power compressors. Centrifugal compressors are typically used due to their ability to handle large volumes of data, which translates to high energy consumption. Turbine-compressor systems are highly reliable and play a critical role in plant operations.

In the mentioned setup, the turbine is commonly powered by high-pressure steam (HPS) which in turn drives the compressor. Hydraulic governor is responsible for regulating the flow of steam to the turbine section. Meanwhile, the electronic governor control panel employs a PID controller to control the opening of the governor.

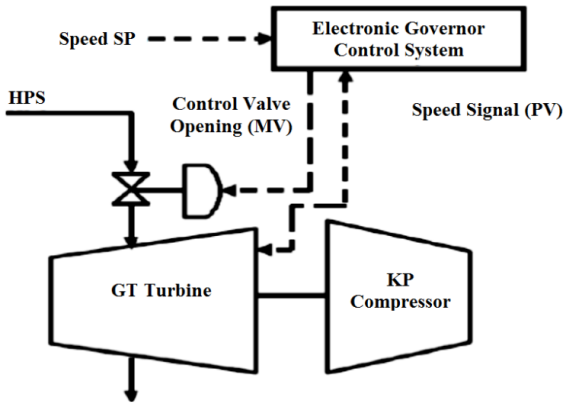


Fig 1. Schematic Diagram of turbine speed Control [10]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:34:13 UTC from IEEE Xplore. Restrictions apply.

As it is widely known, PID controllers are extensively used in various aspects of industrial automation across different industries, including heavy ones like refineries and shipbuilding. However, to optimize the performance, we need to determine the optimal tuning method. Should we still rely on traditional methods like the Ziegler-Nichols technique or move to advanced methods?

II. PID CONTROLLER

P- Proportional, I- Action, and D- Action are the three types of PID controllers also known as ZN PID tune parameters. These controllers are the most widely used control algorithms and will be the focus of this chapter. Feedback loops typically use PID controllers, which come in a variety of configurations such as standalone controllers, part of a DDC package, that is commonly used in plants like pharmaceutical and oil refineries.

It is noteworthy that greater than half of today's industrial controllers are PID controllers. The PID controller is illustrated in the diagram below, and it can be configured in a non-interacting form (NIF).

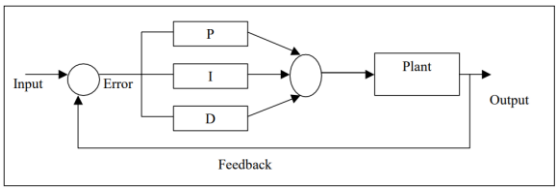


Fig 2. The PID Controller – Non-Interacting Form Schematic Diagram [13]

For proportional control,

$$P = Kp \times e(t)$$

To control the system, it makes use of the proportion of system errors. Offset is produced into the system as a result of this activity.

In Integral control,

$$I = Ki \times \int_0^t e(t)\,dt$$

Its magnitude is proportional to the system's level of inaccuracy. Integral action causes a lag in the system during the action. Offset produced earlier by the Proportional action will be removed as a result of this.

In Derivative control,

$$D = Kd \times d\{e(t)\}/dt$$

It is proportional to error rate of change. Error between desired input and actual output at a specific time t is represented by e(t). By adjusting the values K_p, K_i, and K_d of the PID, various response curves can be generated from a given plant. The challenge in designing a PID for a specific plant is to adjust these values to achieve the desired system performance. It is crucial to keep both the magnitude and duration of the response under an acceptable limit since the response of a system is important. The IAE, can be expressed as:

$$f_{iae} = \int_0^\infty |e(t)|dt$$

Derivative action introduces a lead into the system in the action. This eliminates lag in the system caused by the Integral action earlier in the process.

A. CONTINUOUS PID

The PID controller transfer function:

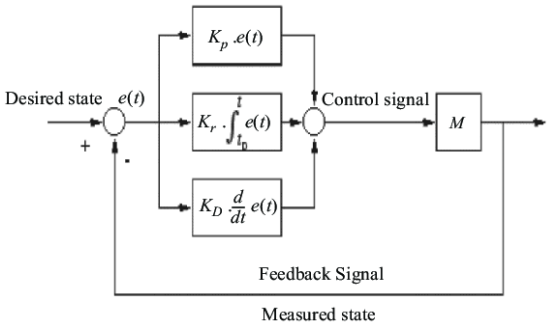


Fig 3. Continuous PID Controller Block Diagram [14]

PID utilizes a combination of three actions to regulate the variable being managed.

The P action is proportional and based on the difference between the feedback and input signals. The Integral action is proportional to the integral of the error signal, while the Derivative action is proportional to the derivative of the error signal.

PID controllers are widely used in industries such as manufacturing, robotics, and process control to regulate variables such as temperature, pressure, flow, and position. They are also used in HVAC systems, automotive control, and flight control systems.

B. DESIGNING PID

The PID controller transfer function is given as

$$G_c(s) = K_p + (K_i/s) + Kds$$

The PID controller's P, I, and D gains are given by K_p, K_i, and K_d, respectively. As depicted in Figure 1, the ACO algorithm is utilized to generate the PID controller Gc(t)'s K_p, K_i, and K_d gains for a specific plant Gp.

$$U(t) = K_p \times e(t) + K_i \int_0^t e(\tau)\,d\tau + K_d \times d\{e(t)\}/dt$$

III. ZIEGLER-NICHOLS METHOD

Ziegler-Nichols introduced the method that requires the unit-step response curve to have an S-shape; parameters like delay time (L) and time constant (T) are obtained by analysing the unit step response of the system, as given in the figure.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:34:13 UTC from IEEE Xplore. Restrictions apply.

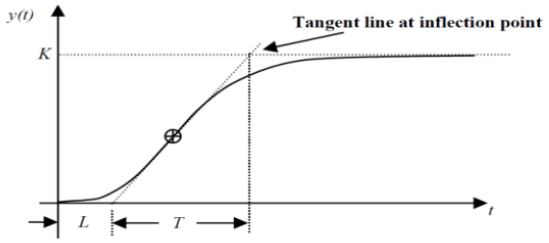


Fig 4. Open Loop System Step Response [12]

Initially, the maximum slope of the response is identified, and a tangent is drawn at the point. The L parameter is obtained by determining the intersection between the tangent and the time axis, while the intersection between the tangent and the line y(t) = K yields T. Parameters are then calculated using the values of a = KL / T and L, as detailed in Table 1.

Controller	Kp	Ti	Td
P	1/*a*	-	-
PI	0.9/*a*	3*L*	-
PID	1.2/*a*	2*L*	*L*/2

Table 1. Ziegler-Nichols Tuning Formula for Step Response Method

A. MANUAL TRAINING METHOD (S-DOMAIN)

Table 2 gives a summary of separate impacts that given three terms of the PID have on the performance of the closed-loop system. It can be utilised to manually adjust the parameters of the controller, starting from the values acquired through ZN methods, and by understanding the impact of the parameters on unit step response. Although the manual tuning is simple, it necessitates some expertise and takes some time to achieve the desired response.

Controller	Rise time T_r	Overshoot $OS\%$	Settling time T_s	Steady-state error ess
Increase P gain [K_p]	Decrease	Increase	Small increase	Decrease
Increase I gain [K_i]	Small decrease	Increase	Increase	Eliminate
Increase D gain [K_d]	Small increase	Decrease	Decrease	Minor change

Table 2. Independent P, I and D effect on the System Response

B. THEORETICAL CALCULATION OF PID PARAMETERS

We further elaborate on PID tuning techniques discussed; a calculation is necessary. In order to do so, let us consider the response of a 3rd order system with a transfer function given by

$$\mathbf{G_p(s)} = \frac{\mathbf{1}}{\mathbf{(3s+1)(2s+1)(s+1)}}$$

To simulate the control system in the s-domain, we will be using Simulink models as depicted in Figure 5.

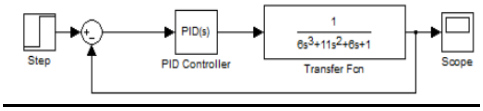


Fig 5. Simulink Model

C. ZIEGLER-NICHOLS STEP RESPONSE CALCULATION

In order to utilize Ziegler-Nichols' calculation method, the step response of the plant must be graphed, as illustrated in the Figure. Response of the system in the figure indicates L = 1.69 T = 8.73, and K = 1.

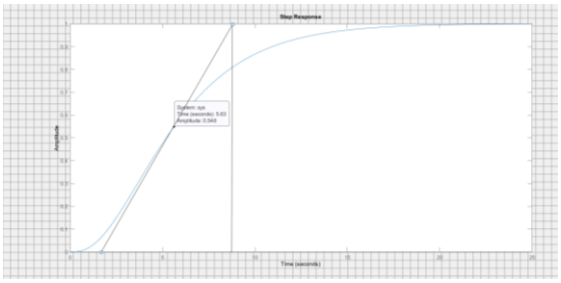


Fig 6. Step Response of Plant Transfer Function

IV. GENETIC ALGORITHM

Genetic Algorithm (GA) is a global search optimization method inspired by natural selection theory. The GA is highly effective at finding solutions to a variety of problems.[6]

In each iteration, the algorithm evaluates the fitness of each solution, and selects the best ones to reproduce and generate new candidate solutions for the next iteration.

They are particularly useful for problems with a large search space, non-linearity, and non-differentiability.

A. POPULATION SIZE

Determining the number of individuals in a population is a crucial step in GA.

The population estimation method used in this study was arbitrary, but based on previous research, a safe range for population size is between 30 and 100 individuals. However, when a sample population of 20 was used in this experiment, the results were discouraging, prompting further testing with populations of 40, 60, 80, and 90 individuals. Ultimately, a population size of 80 was deemed appropriate, with no additional optimization observed beyond a population size of 90 or more.

B. REPRODUCTION

The reproduction step in a genetic algorithm is a crucial process where the fittest individuals from the current population are selected for producing the next generation of solutions. The selection process is based on the fitness function, which evaluates the quality of the solutions. The goal of the reproduction step is to create a diverse population that retains the best solutions from the current population

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:34:13 UTC from IEEE Xplore. Restrictions apply.

while exploring new areas of the solution space.

C. *CROSSOVER*

The crossover operator simulates the genetic recombination process in biological organisms, where the genetic material of two parents is combined to produce offspring with a new set of traits [8].

In the crossover process, two parent individuals are randomly selected from the population and their genetic material is exchanged to produce a new offspring solution. The exchange of genetic material can happen at one or more points along the chromosomes of the parents, and the crossover points are randomly selected for each pair of parents.

The result of the crossover operation is two new offspring solutions that inherit genetic traits from both parents. The new solutions are then evaluated by the fitness function to determine their suitability for the next generation of solutions.

D. MUTATION

The mutation operator introduces random changes to the genetic material of an individual, simulating the process of genetic mutation in biological organisms.

For example, in a binary-coded genetic algorithm, a mutation might involve flipping a bit from 0 to 1 or from 1 to 0. In a real-coded genetic algorithm, a mutation might involve adding a small random value to a gene value.

The mutation operator helps to maintain genetic diversity in the population and prevents the algorithm from getting stuck in local optima. However, the mutation rate should be carefully chosen to balance the exploration and exploitation of the solution space.

V. ANT COLONY OPTIMIZATION

ACO, or Ant Colony Optimization, is an evolutionary meta-heuristic technique that utilizes graphs to solve complex combinatorial optimization problems [5]. The approach involves searching for the lowest cost path through the problem graph, with the help of artificial ants. However, since each ant has limited capabilities, it can only identify suboptimal paths.

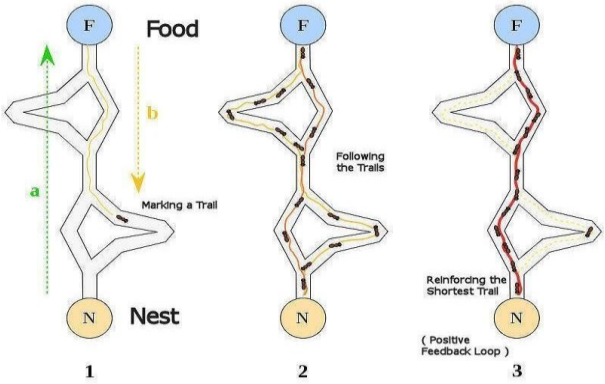


Fig 7. Ant Colony Optimization [11]

Artificial ants in ACO mimic real ants by leaving pheromone trails on the edges of the graph, and choosing their path based on probabilities that depend on the pheromone concentration, which decreases over time. The ants are attracted to nodes with shorter edges and higher pheromone concentration. However, there are also some differences between artificial ants and real ants. Artificial ants exist in a separate universe and move between nodes, while their actions are stored in data structures for memory. Pheromone is usually not deposited during movement, and the amount deposited depends on the quality of the path.

At each generation, artificial ants generate a complete tour by probabilistically selecting nodes in a specific order. In selecting a node, an ant considers both a heuristic factor and a pheromone factor.[1] The heuristic factor, represented by η_{ij}, evaluates the desirability of placing node j at node i in the permutation, and the pheromone factor, represented by r_{ij}, is a measure of the previous successful solutions from other ants. The heuristic value is based on specific problem heuristics, while the pheromone value is based on past successful solutions.[9]

An ant selects the next node using the Pseudo-Random-Proportional Action Choice Rule. This rule states that the ant chooses a node j from the set S of nodes that have not been selected yet based on a probability q_o.[2] The value of qo is a parameter of the algorithm and falls within the range of 0 to less than 1. The ant chooses the node that maximizes a certain criterion, although the specific criterion is not stated.

$$[rij]^{\alpha}[\eta ij]^{\beta}$$

An ant's decision-making process for selecting the subsequent node is influenced by the pheromone and heuristic values, which are determined by the constants α and β. The algorithm employs the Pseudo Random-Proportional Action Choice Rule, and the probability parameter qo is utilized. When qo falls within the range of 0 to less than 1, the ant picks node j from the set S that provides the highest product of pheromone and heuristic values. In contrast, when qo is outside this range, the ant employs a formula to choose the next node based on a probability distribution.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:34:13 UTC from IEEE Xplore. Restrictions apply.

$$P_{ij} = \frac{[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{h \epsilon S}[\tau_{ih}]^{\alpha}[\eta_{ih}]^{\beta}}$$

The likelihood of transitioning between nodes is determined by finding a balance between the heuristic and pheromone factors. Nodes that are closer (i.e., with a lower cost path) should be selected more frequently to implement a greedy constructive heuristic for the heuristic factor. Regarding the pheromone factor, if a particular edge (i, j) has been frequently utilized, it is deemed highly desirable, leading to an autocatalytic process. When the selection of the next node yields a value of P_{ij}, the previously mentioned rule is applied, and the heuristic factor ij is calculated using the following formula.

$$\eta_{ij} = \frac{1}{f(X_j)}$$

In the given equation, f(X) denotes the cost function of X. [4] The above formula prefers choosing edges that are shorter (i.e., lower cost) and have a greater quantity of pheromone. When the ant travels from node i to node j, it leaves a substance on the path, defined as follows:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k} & if\ \text{k}^{\text{th}}\ ant\ uses\ edge\ (i,j)\ in\ its\ tour \\ 0 & otherwise \end{cases}$$

The amount of pheromone laid by ants is related to the quality of the path they take, with more pheromone being deposited on shorter paths. This value is evaluated after an ant completes a tour, which consists of n iterations or generations. The amount of pheromone previously deposited on the trail is then updated based on certain guidelines. The update is done by considering a constant Q and the cost of the tour performed by the ant, denoted by L_k.

$$\tau_{ij}\ (t+n) = \rho.\tau_{ij} + \Delta\tau_{ij}(t)$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^{m} \tau_{ij}(t)$$

where m represents the number of ants, the coefficient (0,1) determines how long the trail persists during a cycle, where (1- ρ) represents the amount of trail that evaporates between generation η_g and η_{g+1}.

The method for modifying pheromone levels aims to replicate the impact of fresh pheromone deposition by ants on the edges visited, as well as the dissipation of pheromone. The process concludes either when an ant finds a solution or the maximum number of iterations is attained.

VI. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is a bio-inspired technique utilized for finding the best possible solutions within a provided solution space. It differs from other optimization algorithms since it solely depends on the objective function and does not rely on the gradient or any differential form of the objective.

The idea behind the algorithm is based on the behaviour of flocks of birds or schools of fish, where all members of the group can benefit from the experiences of others.[7] Essentially, if one bird in a flock finds food, the others can follow and benefit from this discovery.

In the context of optimization, PSO can be thought of as a group of particles moving through a high-dimensional solution space, each particle representing a potential solution. The particles move towards the best solution found so far by the group, while also exploring new areas of the solution space. While it is unlikely that PSO will find the true global optimal solution, it often produces results that are very close to it.

A. OPTIMIZATION PROBLEM

The optimization problem involves a vector variable $X = [x_1x_2x_3...x_n]$, which is used to minimize or maximize the cost function f(X), depending on the optimization formulation. X is also referred to as the position vector, representing a variable model with n dimensions where n is the number of variables in the problem. For instance, in the problem of a flock of birds choosing a landing point, X can be thought of as the latitude and longitude. The fitness function or objective function, f(X), evaluates how good or bad a position X is based on how well a bird can land there. This evaluation is done using multiple survival criteria.

B. PSO ALGORITHM

The following is the step-by-step procedure for the algorithm:

1. Generate a 'population' of agents (particles) that are evenly spread out across the range of X.

2. Compute the value of the objective function for each particle by considering its position. For instance, consider the following function.

$$z = f(x, y) = sinx^2 + siny^2 + sinx\ siny$$

3. If a particle's current position is superior to its previous best position, then replace it.

4. Determine the best particle based on their previous best positions.

5. Modify the velocities of the particles.

$$V_i^{t+1} = W.V_i^t + c_1U_1^t\left(P_{b_1}^t - P_i^t\right) + c_2U_2^t\left(g_b^t - P_i^t\right)$$

6. Update the positions of the particles to their new locations.

$$P_i^{t+1} = P_i^t + v_i^{t+1}$$

7. Continue with the steps described in 2 until the stopping conditions are met.[3]

VII. RESULTS

In this section, we present the results of our proposed algorithms using both numerical and graphical representations. All computations were performed using

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:34:13 UTC from IEEE Xplore. Restrictions apply.

MATLAB/Simulink. In order to compare the performance of our algorithms, we also used a Genetic Algorithm, ACO and PSO. The plant used in our experiments is:

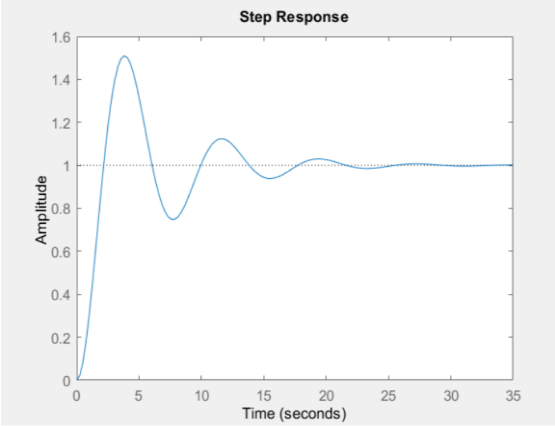


Fig 8: Step-response of PID plant using ZN

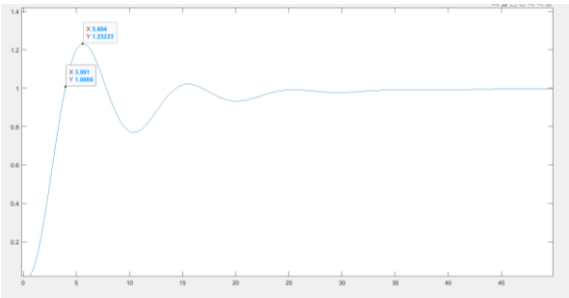


Fig 9: Step-response of PID plant using GA

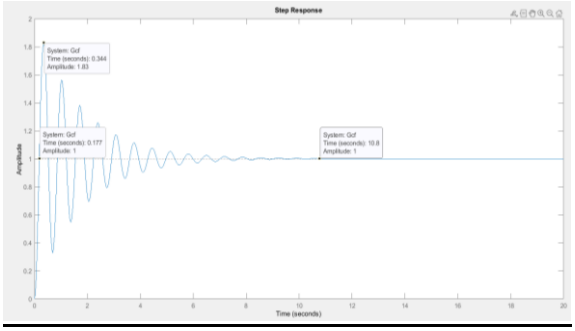


Fig 10: Step-response of PID plant using ACO

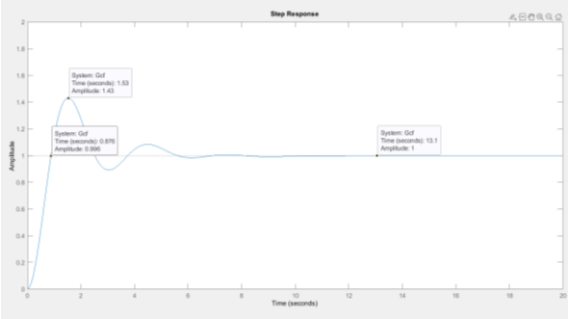


Fig 11: Step-response of PID plant using PSO

Measuring Factors	GA Controller	PSO Controller	ACO Controller	ZN Controller
Rise Time (sec)	4	0.878	0.177	2.15
Maximum Overshoot (%)	23	43	83	51
Settling Time (sec)	35	13.1	10.8	28.2

Table 3. Comparison of different Optimization techniques

Table 3 presents the simulation outcomes of three advanced optimization techniques and one classical method applied to a single plant. The results are based on three measurement criteria. It is evident that all advanced techniques have performed better than the classical method. The classical method performed better only in the Maximum Overshoot column. Nevertheless, in most industrial plants, a slight overshoot is not a major concern. The ACO controller had a much shorter rise time than the GA and PSO controllers, and its settling time was also shorter than that of the GA and PSO controllers. However, the maximum overshoot of the ACO controller was more than double that of the GA controller. We can improve the performance by decreasing the overshoot. Therefore, the ACO controller performed better than the GA and PSO controllers in terms of rise-time and settling-time parameters, but it lagged in the maximum overshoot parameter. Settling time is low in all the metaheuristic techniques, hence the stability is achieved.

IX. CONCLUSION

The study demonstrated the potential of current metaheuristic methods and computational capabilities in obtaining optimal adaptive control. It presented a new method for creating PID controllers that would provide desirable performance for a particular plant. The process of creating these controllers was framed as an optimization problem using four performance indexes, including maximum overshoot, settling time, rise time, and integral absolute error. To identify the most suitable solution, optimization algorithms utilizing GA, PSO, and ACO were created, which enabled the discovery of the optimal global solution.

The unique aspect of the techniques introduced in this paper is that they are easy to implement, provide greater flexibility in problem formulation, and enable designers to obtain a global optimum solution for designing PID controllers. The proposed methods were evaluated on a plant and produced promising results, which were compared to each other as well as a classical method.

X. REFERENCES

[1] Ying-Tung Hsiao, Cheng-Long Chuang, ChengChih Chien. "Ant colony optimization for designing of PID controllers", 2004 IEEE International Conference on Robotics and Automation ;(IEEE Cat. No.04CH37508), 2004

[2] Gustavo Maia de Almeida, Marco Antonio de S. L. Cuadro, Rogerio Passos P. Amaral, Jose Leandro Felix

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:34:13 UTC from IEEE Xplore. Restrictions apply.

Salles. "Optimal tuning parameters of the dynamic matrix predictive controller with ant colony optimization", 2014 11th IEEE/IAS International Conference on Industry Applications, 2014

[3] Houssam Nahilia, Mohamed Boudour. "Distribution power loss minimization using Particle Swarm Optimization and genetic algorithms: Application on Algerian isolated grid", 2016 International Conference on Control, Decision and Information Technologies (CoDIT), 2016

[4] Ying-Tung Hsiao, Cheng-Long Chuang, ChengChih Chien. "Ant colony optimization for best path planning", IEEE International Symposium on Communications and Information Technology, 2004. ISCIT 2004., 2004

[5] Seckiner, S.U. "Ant colony optimization for the job rotation scheduling problem", Applied Mathematics and Computation, 20080715

[6] D.S. Pereira, J.O.P. Pinto. "Genetic algorithm based system identification and PID tuning for optimum adaptive control", Proceedings, 2005 IEEE/ASME International Conference on Advanced Intelligent Mechatronics., 2005

[7] "Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011", Springer Science and Business Media LLC, 2012

[8] Altinten, A.. "Generalized predictive control applied to a pH neutralization process", Computers and Chemical Engineering, 200710

[9] Cheng-Chih Chien. "Ant colony optimization for designing of PID controllers", 2004 IEEE International Conference on Robotics and Automation (IEEE Cat No 04CH37508) CACSD04, 2004

[10] M. V. Subramanyam, K. S. Prasad, and P. G. K. Rao, "Robust Control of Steam Turbine System Speed using Improved IMC Tuned PID Controller," Procedia Engineering, vol. 38, pp. 1450-1456, 2012.

[11] Datta, A & Nandakumar, Soundharyaa. (2017). A survey on bio inspired meta heuristic-based clustering protocols for wireless sensor networks. IOP Conference Series: Materials Science and Engineering. 263. 052026. 10.1088/1757-899X/263/5/052026.

[12] Dinakin, Demilade & Oluseyi, Peter. (2019). Comparative study of meta-heuristic and classical techniques for automatic generation control of a multi-source single-area hydro-thermal-gas power system. 10. 151-172. 10.30572/2018/kje/100310.

[13] Bharadwaj, C. & Thanikanti, Sudhakar Babu & Natarajan, Rajasekar. (2018). Tuning PID Controller for Inverted Pendulum Using Genetic Algorithm. 10.1007/978-981-10-4762-6_38.

[14] Bansal, Hari. (2009). Tuning of PID Controllers using Simulink. International Journal of Mathematical Modeling, Simulation and Applications.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:34:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

Unbalanced Half-bridge Split Capacitor Power Decoupling with Multi-Order Frequency Control for 800V On-board Battery Charger

Tuyen D. Nguyen, *Senior Member, IEEE*, Long H. B. Nguyen, *Student Member, IEEE*, Dong Tran, *Student Member, IEEE*, Ravi Nath Tripathi, *Senior Member, IEEE*, Hai N. Tran, *Member, IEEE*

***Abstract*- This paper proposes a new method for the design and control of a Half-bridge Split Capacitor based Power Decoupling (HSC-PD) circuit to eliminate the bulky Electrolytic Capacitors (E-cap) at dc-link, targeting 800V OBC applications with higher reliability and power density. In the proposed method, both capacitance and dc-offset voltage of the decoupling capacitors are designed unbalanced with the bottom capacitor having a dominant effect over the top one, which reduces the decoupling capacitance requirements compared to the previous works. In addition, the effect of lower dc-link capacitance in reducing power rating of the pre-charge resistor, typically used in OBC applications, is presented. This effect, overlooked in previous literature, leads to reducing volume and cost of the OBC. A small signal model of the HSC-PD is introduced, and a modified feedback control strategy is proposed, offering lower memory usage and calculation effort for the DSP microcontroller when compared with conventional Proportional Resonant (PR) controllers. Comparisons of the proposed method with other HSC-PD counterparts as well as with the conventional E-cap method are given. Finally, a 7.4kW prototype of an 820V interleaved totem-pole PFC with HSC-PD is built and tested to verify the superior performance of the proposed method.**

Index Terms – OBC; power decoupling; electrolytic capacitor; 800V electric vehicles, half-bridge split capacitor.

I. Introduction

Battery Electric Vehicles (BEVs) has been growing rapidly in recent years, and expected to replace the Internal Combustion Engine Vehicles (ICEVs) in the future to reduce CO_2 emission through transportation [1] – [4]. On-board Battery Charger (OBC) is one of the pivotal component of the BEVs, and it's main function is to charge the traction battery at home or office using the available ac grid [5] – [6].

Manuscript received 16 April 2023; revised 21 May 2023; accepted 17 July 2023. This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number DN2022-20-04. (*Corresponding author: Hai N. Tran*).

Tuyen D. Nguyen, Dong Tran are with Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, Ho Chi Minh City, Vietnam. (e-mail: *ndtuyen@hcmut.edu.vn*, *1652154@hcmut.edu.vn*).

Ravi Nath Tripathi is with Kyushu Institute of Technology, 808-0196 Kitakyushu, Fukuoka, Japan. (e-mail: *tripathi.ravi-nath639@mail.kyutech.jp*).

Hai N. Tran, Long H. B. Nguyen are with xEVtech Co. Ltd., Ba Ria city, Ba Ria - Vung Tau, Vietnam. (e-mail: *haitran@xevtech.vn*; *longnguyen@xevtech.vn*).

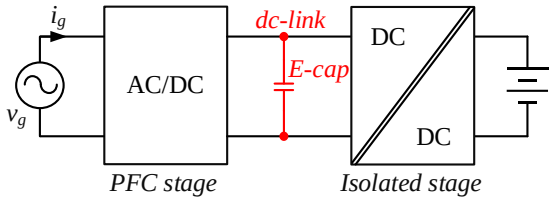


Fig. 1. Conventional OBC with E-cap.

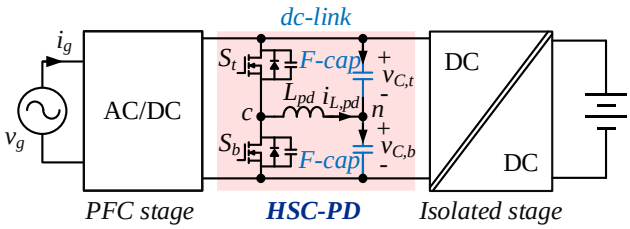


Fig. 2. Proposed E-capless 800V OBC based on *HSC-PD* circuit.

The conventional single-phase OBC shown in Fig. 1 have been widely used in 400V BEVs, in which a large numbers of parallel Electrolytic capacitors are placed between a PFC and an isolated stages to absorb the double-line frequency ripple [7] – [10]. However, these E-caps are short life-time, low current rating, and low power density making it not suitable for the next generation OBC where the requirement of reliability and power density of EV power converters are increased aggressively along with the increasing of charger power rating [5], [11], [12]. In addition, the EV's battery voltage are tend to increase to 800V from current 400V system, leading to many series and parrallel connections of E-cap are required to increase the dc-link votlage in the conventional OBC design method due to low avaiability of high-voltage rating E-cap. This worsens the reliability and power density of the converter. It worth to mention that the use of 800V battery allows to develop a dc fast-charger with higher power rating up to 350kW compared to a typical 50kW-to-100kW fast charger currently available for the 400V system, which significantly reduce charging time. Moreover, various benefits of 800V system over the 400V system such as lighter vehical weight, wider speed range, and higher efficiency have been mentioned in previous literatures [13].

A well-known method to eliminate the E-cap is by integrating an auxiliary Power Decoupling (PD) circuit into the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

main power converter. The PD circuit makes the instantaneous power between ac-grid and dc battery sides balanced. Hence, the capacitance requirement for the dc-link is significantly reduced so that the E-cap can be replaced by film or ceramic capacitors [14] - [25]. The auxiliary storage elements of the PD circuit can be inductor, capacitor, or both. However, capacitor is preferable for PD purposes due to lower loss and higher energy density compared to that of the inductor. This paper proposes a new structure for 800V OBC based on PD circuit, as shown in Fig. 2. A new design approach and new control method for the PD circuit are proposed to reduce the required capacitance and improve its performance, power density and efficiency. In addition, the precharge resistor of the OBC is significantly reduced as a result of dc-link capacitance reduction. Furthermore, a modified feedback control based on co-ordinate transformation *αβ/dq* is proposed to reduce the digital control complexity.

The PD circuits in the literatures is classified in to two methods, AC PD and DC PD. In AC PD method [19] – [21], the voltage across decoupling capacitors is equal to ac-grid voltage. Therefore, high capacitance is required due to zero dc-offset voltage. This high capacitance issue of AC PD method is worsen in OBC applications which usually requires the converter to operate with universal grid voltage, typically ranging from $85V_{ac,rms}$ to $265V_{ac,rms}$, thus, higher capacitance is required considering at the lower grid voltage situation. In DC PD methods [14] – [18], the voltage across decoupling capacitor consists of both ac and dc components with the maximum value up to dc-link voltage which is always higher than grid voltage in the boost PFC converter. Therefore, the required decoupling capacitance in the DC PD method is lower than that of the AC PD method due to high dc-offset component, especially in 800V OBC. It's worth noting that even though the voltage rating of decoupling capacitors in AC PD is lower than DC PD due to zero dc-offset, there is not big different in the actual voltage rating of decoupling capacitors between two motheds because the dc-volage rating of film capacitors is usually much higher than its ac-voltage rating [26], [27]. Among various DC PD circuit topologies were presented in the previous works [14] – [18], Half-bridge Split Capacitor based PD (HSC-PD) is frequently used because the dc-link capacitors are directly utilized for power decoupling purpose, thus, no addition capactior is needed [15] – [18].

Several control methods have been developed for the HSC-PD converter [15] – [18]. In [16], the HSC-PD structure was first introduced with the open-loop control method. However, open-loop control is sensitive to system uncertainties, such as parameter tolerances and disturbance, leading to not accuracy. After that, a closed-loop control algorithm for the HSC-PD circuit is proposed in [18], which is much more robust under various uncertainty and steep load change conditions. In both [16] and [18], decoupling capacitors are designed in such a way that their capacitance value and dc voltage component are equal, and their ac voltage component fluctuates at a line frequency. However, this strategies have several drawbacks: 1) large capacitance value of both decouling capacitor, which increases not only capacitor volume but the current rating of the PD circuit; 2) difficult to select the decoupling capacitor since its ac voltage rating is high; 3) low efficiecny because of high current rating and equivalent step-down ratio; 4) low dynamic response since controlling the ac components at line-frequency requires a 90^o-shifting function in the controller.

In [17], a unequal dc voltage component method is introduced in a complete closed-loop control system. This reduces the amplitude of ac voltage component. Moreover, the ac component is controlled at the double-line frequency to avoid the delay in the controller and make it insensitive with the capacitance tolerance. Even though the robustness of this control method has been practically demonstrated under large capacitance tolerance conditions, it is designed based on an equal capacitance approach, resulting in large equivalent capacitance and high current rating. In addition, the effect of unequal dc voltage component on converter efficiency, and the fourth-line frequency harmonic caused by double-line frequency voltage and current in the decoupling circuit have not been discussed.

In [15], unequal capacitance approach was first introduced to improve the robustness of the controller under non-ideal grid, and the ac voltage ripple is obtained with multi-harmonic frequency by using conventional PR controller. However, the ac voltage rating of capacitors is high due to equal dc voltage component, making it not suitable for high dc-link voltage converter such as 800V OBC since the high ac voltage rating capacitor is usually bulky and low avaiability. In addition, the capacitance values in [15] are pretty high of 330μF and 120μF for the top and bottom capacitors, respectively, at 400V dc-link and 1.5kW.

Considering the aforementioned issues, this paper proposes a new design approach and control method for the HSC-PD circuit, in which both dc component and capacitance values of the decoupling capacitors are designed unequally with bottom-bias capacitor approach, reducing not only equivalent capacitance and ac voltage ripple but current rating of the PD circuit. The effect of fourth-line frequency harmonics, which is caused by double-line frequency of voltage and current in the PD circuit, is also analyzed. The dc-offset ratio and its impact on the converter efficiency will be discussed. Furthermore, the feedback control of HSC-PD is developed for the entire system considering the proposed unequal dc component and capacitance. The multi-order frequency harmonic component compensated through the PR controller, however, the implementation in digital domain consdiering *αβ* frame requires considerable amount of delay-time for computation as well as frequency response is poor as compared to the integral control action. In this paper, implementation of PR controller is proposed using *αβ-dq* transformation with an integrator that enables the generation of multi-order frequency component, reduction in DSP memory utilization and calculation effort. Theoretical analysis and experimental results are given to verify the superiority of the proposed method.

II. Proposed Even-order Frequency Unbalanced Half-bridge Split Capacitor Power Decoupling Control

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

A. Theoretical analysis and design guideline

Circuit diagram of the proposed 800V OBC is shown in Fig. 2, in which bulky E-cap is replaced by HSC-PD circuit using low capacitance Film capacitor (F-cap). The charger is rated at 7.4kW, and the dc-link voltage is fixed at 820V. Compared to the variable dc-link strategy as presented in [7], the use of fixed dc-link voltage at maximum provides the benefits of: 1) minimize the required decoupling capacitance; 2) minimize the current conducted in the primary side of the isolated stage of the OBC reducing current rating and conduction loss. It is worth mention that the cost of 1200V SiC MOSFET, that is used in the primary side of the isolated stage, increases significantly with its current rating. The grid current and voltage as well as its instantaneous power can be expressed as below:

$$\begin{cases} v_g(t) = \sqrt{2} \cdot V_g \cdot \sin(\omega \cdot t + \varphi) \\ i_g(t) = \sqrt{2} \cdot I_g \cdot \sin(\omega \cdot t + \varphi) \end{cases} \quad (1)$$

$$p_g(t) = v_g(t) \cdot i_g(t) = V_g \cdot I_g - V_g \cdot I_g . \cos(2 \cdot \omega \cdot t + 2 \cdot \varphi) \quad (2)$$

The phase of grid voltage and current φ will be assumed to be zero to reduce the complexity in the later analysis since power factor correction is used. The first term on the right side of (2) represents the average power charged the battery, P_{dc}, while second term is the second-order power that needs to keep away from the battery. In the conventional OBC, this second-order power is stored in the dc-link using bulky electrolytic capacitors leading to low power density and reliability. In this paper, the second-order power is stored in two small film capacitors C_t and C_b by using the proposed design and control method for the HSC-PD, which is described as follows.

The voltage equation for each decoupling capacitors, $v_{C,t}(t)$ and $v_{C,b}(t)$, are proposed as:

$$\begin{cases} v_{C,t}(t) = V_{dc,t} + \sum_{n=2,4,6,...} V_{ac(n)} \cdot \sin(n \cdot \omega \cdot t + \theta_n) \\ v_{C,b}(t) = V_{dc,b} - \sum_{n=2,4,6,...} V_{ac(n)} \cdot \sin(n \cdot \omega \cdot t + \theta_n) \end{cases} \quad (3)$$

The first and second terms on the right side of (3) represent dc and ac voltage components of each decoupling capacitor respectively. The ac voltage components include multiple of even-order frequencies $V_{ac(n)}$ ($n = 2, 4, 6, \ldots$), and are 180° phase shift between two capacitors to ensure the sum of which are constant at the desired dc-link voltage. The block diagram of the proposed controller is detailed in Section III. Ideally, only second-order ac voltage component is needed for storing the second-order power of the grid in (2). However, fourth-order power is created because of second-order voltage and current conducted through the decoupling capacitors. Similarly, higher even-order powers of 6th, 8th, … are created from its previous even orders. Therefore, the decoupling capacitor voltages must be operated with multi even-order frequencies to suppress the second-order power of the grid and other higher even-order powers created by itself, as shown in (3).

The dc voltage components of two capacitors, $V_{dc,t}$ and $V_{dc,b}$, are unequal with the proposed dc-offset factor m as below:

$$\begin{cases} V_{dc,t} = V_{Link} \cdot \left(\frac{1}{2} - m\right) \\ V_{dc,b} = V_{Link} \cdot \left(\frac{1}{2} + m\right) \end{cases} \quad with\ 0 < m < 0.5 \quad (4)$$

It can be seen from (4) that in the proposed method dc voltage component of the bottom capacitor $V_{dc,b}$ is always kept greater than that of top capacitor, $V_{dc,b} > V_{dc,t}$, since m is the dc-offset ratio, and can be chosen within $0 < m < 0.5$. This proposed approach is different from the previous work presented in [15] that was designed with higher dc voltage component of the top capacitor. The reason for this proposed approach is that if considering the HSP-PD circuit as a buck converter with input and output voltages are V_{Link} and $V_{dc,b}$, respectively, higher $V_{dc,b}$ leads to lower step down ratio of buck converter and narrower duty cycle range. In addition, higher bottom capacitor voltage led to lower decoupling inductor current for the same power rating. Therefore, the proposed unbalancing approach with higher dc voltage component of the bottom capacitor achieves higher efficiency and smaller decoupling inductor as compared to the previous work [15], [17]. It is noticeable that the peak ac voltage component must be smaller than the dc voltage component of the top capacitor to avoid over modulation.

Current conducted through two decoupling capacitors are determined from (3) as below:

$$\begin{cases} i_{C,t}(t) = C_t \cdot \frac{dv_{C,t}(t)}{dt} \\ \qquad = C_t \cdot \sum_{n=2,4,6,...} n \cdot \omega \cdot V_{ac(n)} \cdot \cos(n \cdot \omega \cdot t + \theta_n) \\ i_{C,b}(t) = C_b \cdot \frac{dv_{C,b}(t)}{dt} \\ \qquad = -C_b \cdot \sum_{n=2,4,6,...} n \cdot \omega \cdot V_{ac(n)} \cdot \cos(n \cdot \omega \cdot t + \theta_n) \end{cases} \quad (5)$$

where C_t and C_b are capacitance values of top and bottom capacitors, respectively. Current conducted through decoupling inductor can be obtained from (5) by:

$$\begin{aligned} i_{L,pd}(t) &= i_{C,b}(t) - i_{C,t}(t) \\ &= -(l+1) \cdot C_t \cdot \sum_{n=2,4,6,...} n \cdot \omega \cdot V_{ac(n)} \cdot \cos(n \cdot \omega \cdot t + \theta_n) \end{aligned} \quad (6)$$

Typical waveforms of the proposed control method are shown in Fig. 3.

Not only dc voltage component but also capacitances of two decoupling capacitors are designed unequally in the proposed method. Assuming that l is capacitance ratio between two decoupling capacitors:

$$l = \frac{C_b}{C_t} \quad (7)$$

The instantaneous power stored in two capacitors is therefore obtained from (3) - (6) by:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

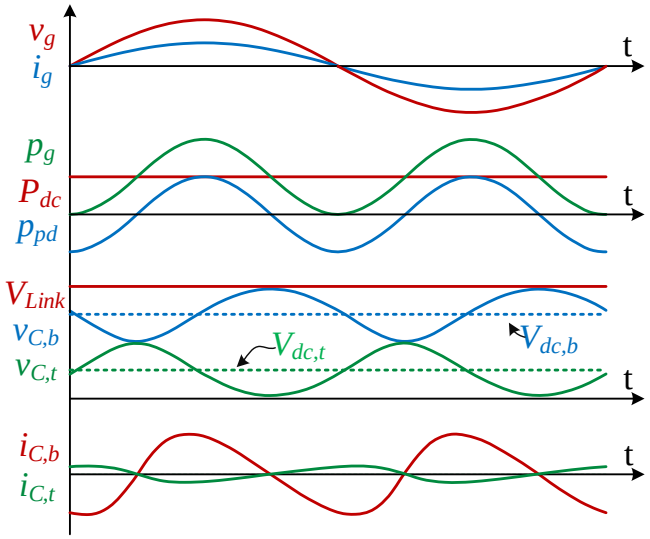


Fig. 3. Typical waveforms of the proposed power decoupling method at full load

$$\begin{aligned} p_{pd}(t) &= p_{C,t}(t) + p_{C,b}(t) \\ &= v_{C,t}(t)\cdot i_{C,t}(t) + v_{C,b}(t)\cdot i_{C,b}(t) \\ &= -\omega\cdot\left(V_{dc,b}\cdot C_b - V_{dc,t}\cdot C_t\right)\cdot \sum_{n=2,4,6,\ldots} n\cdot V_{ac(n)}\cdot\cos(n\cdot\omega\cdot t+\theta_n) \\ &+(l+1)\cdot C_t\cdot \sum_{n=2,4,6,\ldots}\left(V_{ac(n)}\cdot\sin(n\cdot\omega\cdot t+\theta_n)\right)\ldots \\ &\ldots\cdot \sum_{n=2,4,6,\ldots}\left(n\cdot\omega\cdot V_{ac(n)}\cdot\cos(n\cdot\omega\cdot t+\theta_n)\right) \end{aligned} \tag{8}$$

The phase angle θ_n can be determined by:

$$\theta_n = -(n-2)\cdot\frac{\pi}{4} \tag{9}$$

The peak value of each ac voltage components in (8) are determined by:

$$\begin{cases} V_{ac(2)} = \dfrac{V_g\cdot I_g}{\omega\cdot\left[l-1+2\cdot m\cdot(l+1)\right]\cdot C_t\cdot V_{Link}} \\ V_{ac(4)} = \dfrac{(l+1)\cdot V_{ac(2)}^2}{2\cdot\left[l-1+2\cdot m\cdot(l+1)\right]\cdot V_{Link}} \\ V_{ac(6)} = \dfrac{(l+1)\cdot V_{ac(2)}\cdot V_{ac(4)}}{\left[l-1+2\cdot m\cdot(l+1)\right]\cdot V_{Link}} \\ V_{ac(8)} = \dfrac{(l+1)\cdot\left(2\cdot V_{ac(4)}\cdot V_{ac(4)}+4\cdot V_{ac(2)}\cdot V_{ac(6)}\right)}{4\cdot\left[l-1+2\cdot m\cdot(l+1)\right]\cdot V_{Link}} \end{cases} \tag{10}$$

The first term on the right side of (8) represents for the second-order power that is stored in the decoupling capacitors and the rest terms in the right side of (8) are the power naturally generated by the controller itself and is controlled to store within the decoupling circuit. The second order and self-created power components (4th, 6th, 8th, …) with different value of *m* are shown in Fig. 4. Practically, the waveforms of voltage (current) of decoupling capacitor (inductor) is generated by a controller in the closed loop system, thus to reduce the control complexity

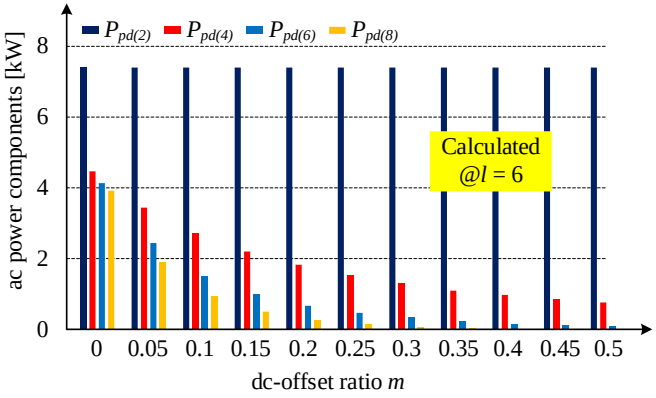


Fig. 4. AC power components of the decoupling circuit calculated at l = 6, C_t = 16 μF, and C_b = 86 μF.

caused by the endless chains of (3) and (6), the orders higher than eight will be ignored since the amplitude of which are negligible. It can also be seen from Fig. 4 that second-order power in the proposed approach is independent from *m* since it is designed same as OBC rating, and the higher dc-offset ratio *m* will lead to lower other self-created ac power in the circuit (higher order ac power of 4th, 6th, 8th, …), this has not been investigated in the previous literatures. If second-order power of the grid in (2) are equal to that from (8) to achieve the ideal decoupling purpose, thus:

$$\left[l-1+2\cdot m\cdot(l+1)\right]\cdot C_t = \frac{V_g\cdot I_g}{\omega\cdot V_{ac(2)}\cdot V_{Link}} \tag{11}$$

Since decoupling capacitors are designed to absorb second-order power of the grid, the calculation of which can be done by setting *m* to zero to reduce the complexity. The relationship of top and bottom capacitors is therefore obtained by:

$$(l-1)\cdot C_t = C_b - C_t = \frac{2\cdot V_g\cdot I_g}{\omega\cdot V_{Link}^2} = \frac{2\cdot P_g}{\omega\cdot V_{Link}^2} \tag{12}$$

The relationship of $C_{bot} - C_{top} \geq 70$ uF is therefore obtained based on (12) with P_g = 7.4kW and V_{Link} = 820V. In the proposed method, the bottom capacitor with higher dc voltage component and capacitance value absorbs a higher amount of second-order power compared to the top capacitor which has lower dc voltage component and capacitance value.

The equivalent capacitance of dc-link as a function of switching frequency and desired ripple is expressed by:

$$C_{eq} = \frac{P_g\cdot D}{V_{Link}\cdot\Delta V_{Link,sw}\cdot f_s} \tag{13}$$

where *D* is a duty cycle of the PFC circuit, $\Delta V_{Link,sw}$ is the dc-link voltage ripple at switching frequency, and f_s is the switching frequency of the PFC circuit.

The above equivalent capacitance is also obtained based on each decoupling capacitor by:

$$C_{eq} = \frac{C_b\cdot C_t}{C_b + C_t} = \frac{\left(C_t + \dfrac{2\cdot P_g}{\omega\cdot V_{Link}^2}\right)\cdot C_t}{2\cdot C_t + \dfrac{2\cdot P_g}{\omega\cdot V_{Link}^2}} \tag{14}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

Therefore, minimum capacitance of top capacitor can be obtained from (13) and (14) by:

$$C_{t,\min} = \sqrt{\left(\frac{P_g}{\omega \cdot V_{Link}^2} - C_{eq}\right)^2 + C_{eq} \cdot \frac{2 \cdot P_g}{\omega \cdot V_{Link}^2}} - \left(\frac{P_g}{\omega \cdot V_{Link}^2} - C_{eq}\right) \tag{15}$$

The required capacitances as a function of switching frequency are shown in Fig. 5. In this project, f_s and $\Delta V_{Link,sw}$ is set to 50kHz and 8V, respectively. The minimum capacitance of top and bottom capacitors are therefore obtained by 16uF and 86uF, respectively, yielding l_{min} = 5.4.

The comparison of decoupling capacitors between the proposed method and previous research is shown in Fig. 6. It can be observed that AC power decoupling has larger capacitance due to zero dc voltage component [19] – [21]. The line-frequency decoupling approach with balancing capacitors has large capacitance since its equivalent capacitance is high, which is calculated same as the right side of (12). Moreover, due to balancing capacitance approach, each capacitor must be twice of the equivalent capacitance worsening the required capacitance value [14] - [17]. Authors in [15] used the unequal capacitance to increase the robustness of controller under nonideal grid. However, its unbalancing strategy is opposite from our proposed strategy in such a way that bottom capacitor in [15] has smaller capacitance value than top capacitor. This strategy is however not beneficial because it reduces the power decoupling capability, as seen in (8). Moreover, equivalent capacitance given in [15] is same as [16] – [18] which is much larger than this paper.

The relationship between peak 2nd harmonic current of the decoupling inductor as a function of *m* and *l* is described in (16), which is obtained from (6) and (10), and is shown in Fig. 7.

$$i_{L,pd(2)} = \frac{-(l+1) \cdot C_t \cdot 2 \cdot V_g \cdot I_g \cdot \cos(2 \cdot \omega \cdot t + \theta_2)}{\left[l - 1 + 2 \cdot m \cdot (l+1)\right] \cdot C_t \cdot V_{Link}} \tag{16}$$

It can be observed that with same decoupling capacitors, the peak current of decoupling inductor reduces as *m* and *l* increase. However, with a given second-order power and dc-link voltage, m must be designed so that the peak voltage of the bottom capacitor does not exceed dc-link voltage. The selection of decoupling capacitors is a challenging task since their voltages consist of both dc and low-frequency ac components. The dc-voltage rating of the decoupling capacitor is selected based on its peak voltage which is assumed to be dc-link voltage as the

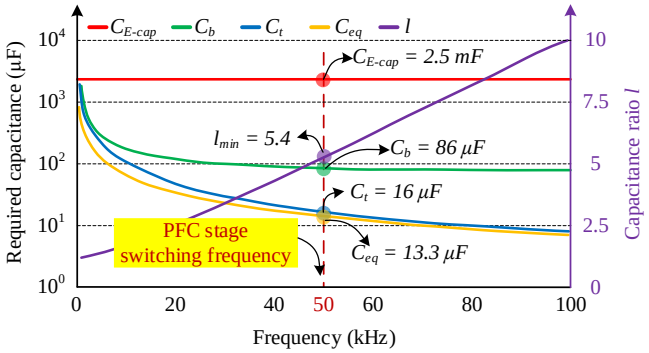


Fig. 5. Required capacitance calculated at $\Delta V_{Link,sw}$ = 8V.

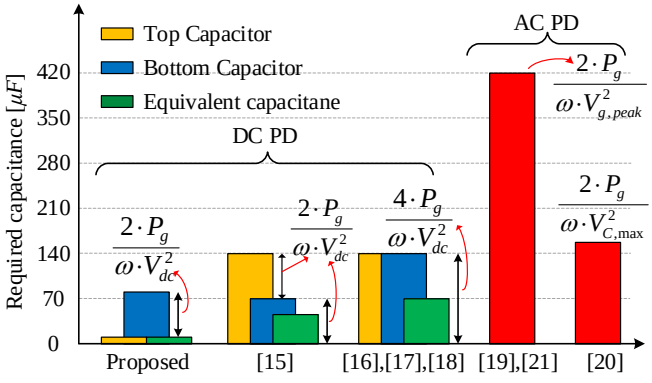


Fig. 6. Capacitance comparisons between the proposed approach and other counterparts

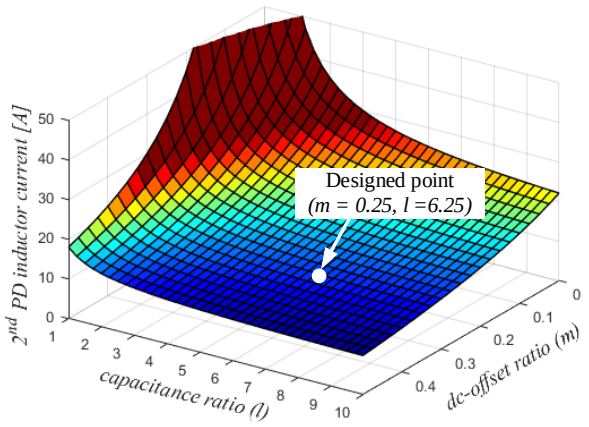


Fig. 7. PD inductor current as a function of *m* and *l*.

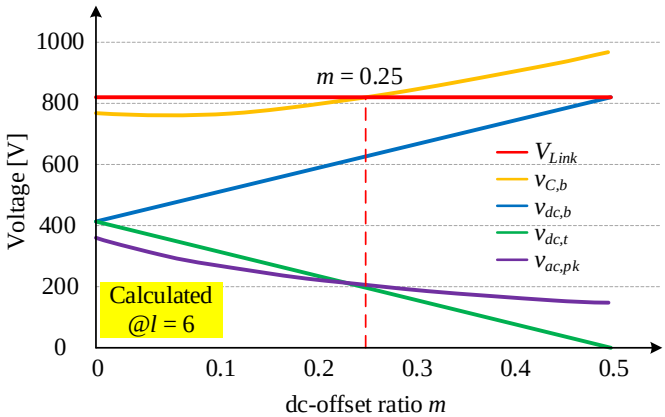


Fig. 8. Decoupling capacitor voltages as a function of *m*

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

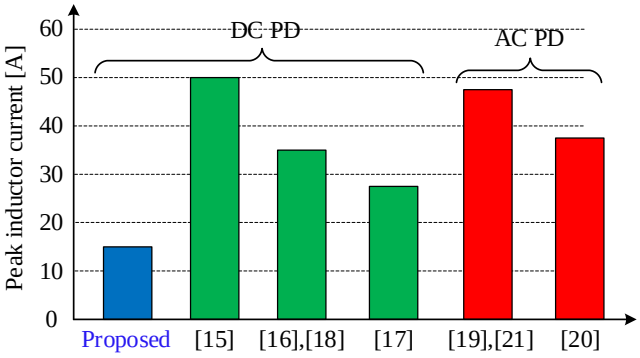


Fig. 9. Decoupling inductor current comparison

worst case in this paper, while ac voltage rating depends on *m*. Lower value of m leads to higher ac voltage rating and vice versa. Since film capacitor is used for decoupling purpose, the dc voltage rating of which is usually several times higher than its ac voltage rating. Fig. 8 shows the dc and ac components of the decoupling capacitors. In this paper, considering the availability and margins, the actual value of decoupling capacitors used for experiment are C_{top} = 15uF and C_{bot} = 100uF, l = 6.25. The dc-link voltage is design maximum at 820V, thus film capacitor BLH505K901A062 with ac voltage rating of 230V_{rms} is selected considering its low-profile benefit. Therefore, *m* can be theoretically designed between 0.05 to 0.25 considering 10% margin for ac voltage rating. However, in practical *m* is designed at its maximum possible value of 0.25 to minimize the self-created power and inductor current, as seen in Fig. 4.

The comparisons of the decoupling inductor current (peak) between the proposed method and other counterparts are shown in Fig. 9. It can be observed that, the proposed method has lowest inductor current due to lower top capacitance value. The unbalance strategy in [15] has another disadvantage is that its inductor current is very high due to higher top capacitance value.

III. Modelling and Control Algorithm

This section describes the proposed control algorithm for the system including PFC stage and HSC-PD circuit. A small-signal model of HSC-PD will be presented, after that, the implementation of two parallel controllers is explained to realize the best compromise between the PFC stage and the PD circuit.

1) Modeling HSC-PD circuit

HSC-PD circuit in Fig. 2 can be considered as a buck converter with input voltage and output voltage are V_{Link} and $v_{C,b}$, respectively. Hence, the voltage across top capacitor $v_{C,t}$ will be naturally obtained as a result of these two input and output voltages. Top switch S_t is the main switch of the converter, and its duty cycle is controlled to generate the desired output voltage, while the bottom switch S_b is switched complementarily with S_t. The converter operates under continuous conduction mode. The average model of HSC-PD with current and voltage is defined as:

$$\begin{cases} \bar{v}_{cp}(t) = d(t)\cdot \bar{v}_{ap}(t) \\ \bar{i}_a(t) = d(t)\cdot \bar{i}_c(t) \end{cases} \tag{17}$$

The average voltage and current are functions of duty cycle '*d*' and termed as the averaged PWM switch model. The linearization of PWM switch model and ignoring second-order terms will result in the small signal PWM switch model under CCM as:

$$\begin{cases} V_{cp} + \hat{v}_{cp}(t) = D\cdot V_{ap} + \hat{d}(t)\cdot V_{ap} + D\cdot \hat{v}_{ap}(t) \\ I_a + \hat{i}_a(t) = D\cdot I_c + \hat{d}(t)\cdot I_c + D\cdot \hat{i}_c(t) \end{cases} \tag{18}$$

The small signal circuit model of the HSC-PD circuit is therefore obtained, as shown in Fig. 10, which is used for the design and development of the controller. The transfer function for duty cycle-to-inductor current (19) and duty cycle-to-output voltage (20) is obtained by considering $\hat{v}_{Link}(s)$ = 0. Here, the output voltage is the ac ripple component in the output dc voltage of HSC-PD circuit. The duty cycle to output voltage transfer function of HSC-PD circuit is similar to [16] and capacitor voltage possessing the same ac component.

$$G_{PDi}(s) = \frac{\hat{i}_{Lpd}(s)}{\hat{d}_{pd}(s)} = \frac{(l+1)\cdot C_t \cdot V_{Link}\cdot s}{L_{pd}\cdot(l+1)\cdot C_t \cdot s^2 + 1} \tag{19}$$

$$G_{PDv}(s) = \frac{\hat{v}_{ac(n)}(s)}{\hat{d}_{pd}(s)} = \frac{V_{Link}}{L_{pd}\cdot(l+1)\cdot C_{top}\cdot s^2 + 1} \tag{20}$$

From (19) and (20), current-to-output voltage transfer function of HSC-PD circuit is obtained as

$$G_{PDiv}(s) = \frac{\hat{v}_{ac(n)}(s)}{\hat{i}_{Lpd}(s)} = \frac{1}{s\cdot(l+1)\cdot C_t} \tag{21}$$

2) Proposed control algorithm

PI-controller is the most commonly used for the dc reference regulation without steady-state error in synchronously rotating dq frame. PR controller are employed for the reference tracking in stationary frame αβ [28]**.** The PI controller in synchronously rotating dq frame is equivalent to the PR controller in stationary αβ frame [29], however, this holds true mainly for the fundamental frequency components. Considering this, the PR controller requires lesser computation burden as there is no transformation required. PI- controller is capable of tracking the dc reference corresponding to the

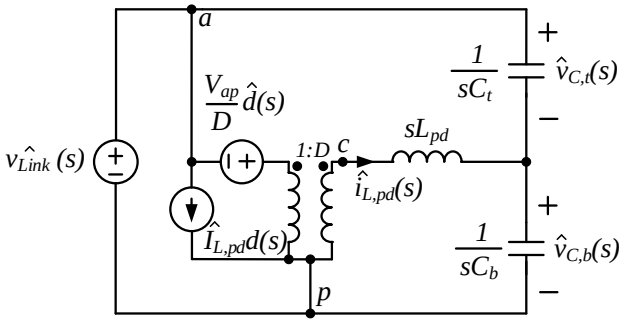


Fig. 10. Small signal model of HSC-PD circuit

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

fundamental freuqncy only. Therefore, the PR controller is required in the feedback control loop for multi-frequency control considering the double-line frequency and fourth line frequency compensation.

The proposed control loop in Fig.11 is the amalgamation of PFC-stage controller and PD controller offering multifrequency compensation with an additional duty-based feedforward control. The top part of the control structure is the PFC stage consisting of a PI-based outer voltage control loop followed by co-ordinate transformation and a PI-based inner-current control loop in a stationary frame. The bottom part of the control structure consists of the PR-based multifrequency control in the rotating synchronous frame and duty-based feedforward control in the inner current control loop. The top and bottom controller structure is employed as an independent parallel control loop with the same bandwidth response for the inner current control loop.

The PFC stage controller responsible for the regulation of dc-link voltage. The dc voltage error fed to PI controller of the outer voltage control loop providing the equivalent dc current. After the co-ordinate transformation, the ac grid current error passed through the PI controller of the inner current control loop to generate reference duty cycle for the switches.

$$G_{PI}(s) = K_p + \frac{K_i}{s} \tag{22}$$

The design of the PFC-stage controller must be in accordance with the HSC-PD controller to achieve the optimized performance of the system. The proposed active power decoupling control technique is intended to reduce the equivalent capacitance. However, the value of capacitance and ripple in the voltage are correlated. Reduction in dc capacitance leads to an increase in the voltage fluctuation at the dc-bus. It is crucial during the transients under load change and change in input power. Therefore, the outer voltage loop of PFC is required to have a higher bandwidth compared to the conventional PFC scheme to regulate dc-link voltage faster through current reference. Considering the above-mentioned issue and criteria, the cut-off frequency of the inner current loop is chosen to be $f_{c_inner} = \frac{f_{sw}}{10}$ and the cut-off frequency of the outer voltage loop is chosen to be at least 2 times of grid frequency $f_{c_outer} \geq f_{grid} * 2$. In [17] and [18] the outer voltage loop cut-off frequency is between 200-300Hz, and higher capacitance values are used. The lower capacitance values used in this paper and the outer voltage control loop is designed with slightly higher bandwidth to accommodate that effect. The frequency response using the bode plot is demonstrated in Fig. 12 for the PFC controller including both the voltage control loop as well as the current control loop. As mentioned above, the cut-off frequency for the outer voltage control loop is $f_{c_outer} = 370Hz$ with 90^0 phase margins. The inner current control loop is designed with cut-off frequency of 7 kHz and 75^0 phase margins.

The multi-frequency decoupling control is implemented with unequal dc offset for the designed HSC-PD circuit having an unequal capacitor. PR controller had been the choice of controller for the multi-frequency control as it possesses the infinite gain at the desired frequency that provides the zero steady-state error. The PR controller transfer function is given by (23) with a damping constant near the resonance frequency for system stability. PR-controller is used to compensate for the multi-frequency harmonic components specially double-line and four-line frequency components in the capacitor voltage. The multi-even-harmonic component is generated using current reference for the inductor in PD circuit.

$$G_{PR}(s) = K_p + \frac{2 \cdot K_r \cdot \omega_c \cdot s}{s^2 + 2 \cdot s \cdot \omega_c + \omega_0^2} \tag{23}$$

Nevertheless, the implementation of the PR controller in the digital environment is an issue because of the programming complexity for a real experimental system. The implementation of the PR controllers on the digital processor requires dicretization and two times delay corresponding to the discretization of the transfer function. The PR-controller is implemented using the second-order generalized integrator (SOGI) in the digital environment [30] possesses complex

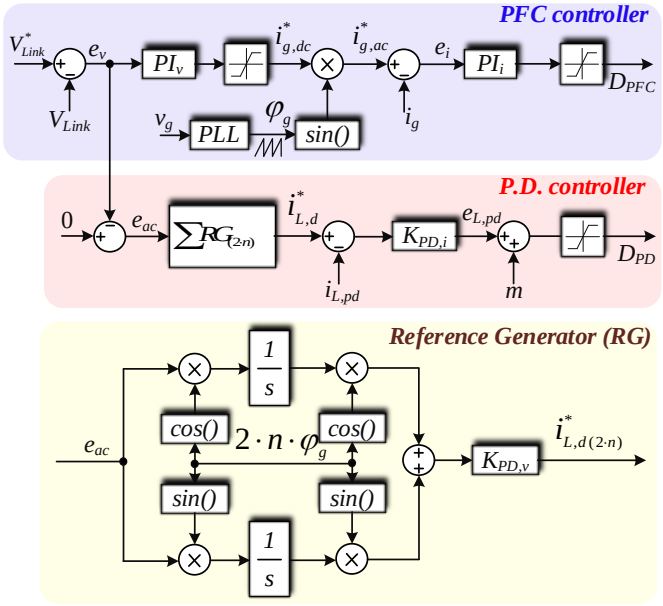


Fig. 11. Block diagram of the proposed control method

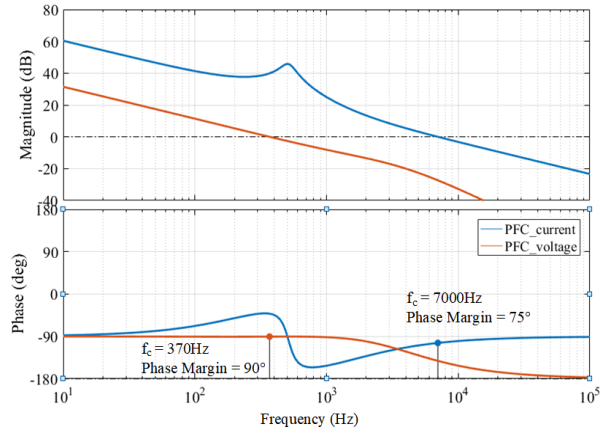


Fig. 12. Bode plot of dual-loop control of the PFC circuit

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

structure and delay. The discretization method significantly impacts the Stability and performance of the PR controller [31]. Therefore, needed a special attention to avoid instability by restraining the actual performace of PR. The frequency response is comparatively not good compared to the integral control action of the PI-controller.

In this paper, the co-ordinate transformation $\alpha\beta/dq$ is proposed in Fig.11 to change n-order-line-frequency (n=2,4,6,..) to DC signal for the implementation of the PR controller. The integral control action is employed to generate equivalent d-axis and q-axis signals and further pass through dq/αβ transformation with infinite DC gain instead of limit gain in [28]. The implementation of the integral control action is simpler and ultimately results in the reduced DSP memory utilization and calculation weights, consequently, making it cost-effective as digital processor cost depends on memory and computing capability. The unbalanced $\alpha\beta/dq$ transformation [32] is used instead of conventional $\alpha\beta/dq$ transformation for easier implementation. The conventional $\alpha\beta/dq$ need 90° delay to α input signal to generate β axis signal through OGM - Orthogonal (imaginary) term generation method. The difference is unbalanced $\alpha\beta/dq$ transformation has a half gain and a negative term that have no significance.

The outer voltage loop of HSC-PD should have to be designed with larger bandwidth as compared to PFC stage outer voltage loop. The outer voltage loop bandwidth of the HSC-PD circuit has a dominant effect on the system performance. The smaller bandwidth compared to the PFC stage lead to a very high voltage overshoot at dc-link. Furthermore, the bandwidth similar to the PFC stage has a better response considering overshoot at dc-link, however, leads to an oscillation in dc voltage and grid current. This ultimately results in the system unstability. Therefore, the larger outer voltage

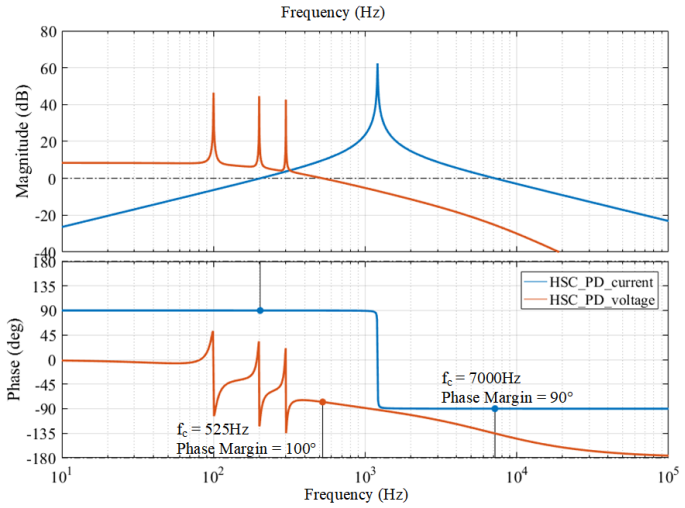
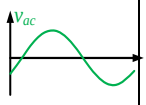
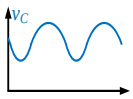
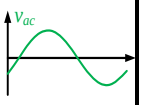
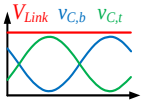
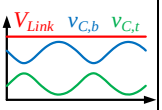
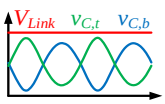
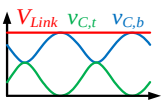


Fig. 13. Frequency response of HSC-PD circuit.

loop bandwidth of the HSC-PD circuit is desirable to achieve a faster response with a minimized overshoot at dc-link voltage. The inner current loop of HSC-PD should have a cut-off frequency at $f_{c_inner} = \frac{f_{sw}}{10}$ and resonant gain should be tuned to have lower response than the PFC stage but still get enough gain to attenuate ripple on DC-Link. The frequency response of HSC-PD circuit in Fig.13 is investigated using bode-plot for current control loop as well as voltage control loop. The current control loop has a cut-off frequency of 7000Hz and 90^0 phase margins. The voltage control loop has three resonant frequencies to compensate the double-line (100Hz), fourth line (200Hz) and sixth (300Hz) frequency harmonics with cut off frequency of 525Hz and 100^0 phase margins. The higher cut-off frequency of outer voltage. Finally, the overall comparison

TABLE I
COMPARISONS OF THE PROPOSED HSC-PD WITH OTHER COUNTERPARTS

AC Decoupling			DC Decoupling			
[19]	[20]	[21]	[16],[18]	[17]	[15]	**Proposed**
						
$C_{pd} = \frac{2 \cdot P_g}{\omega \cdot V_{g,peak}^2}$	$C_{pd} = \frac{2 \cdot P_g}{\omega \cdot \left(\frac{2}{3} \cdot V_{Link}\right)^2}$	$C_{pd} = \frac{2 \cdot P_g}{\omega \cdot V_{g,peak}^2}$	$C_{eq} = \frac{2 \cdot P_g}{\omega \cdot V_{Link}^2}$	$C_{eq} = \frac{2 \cdot P_g}{\omega \cdot V_{Link}^2}$	$C_{eq} = \frac{2 \cdot P_g}{\omega \cdot V_{Link}^2}$	$C_{eq} = \frac{P_g \cdot D}{V_{Link} \cdot \Delta V_{Link,sw} \cdot f_s}$
			$C_{bot} = C_{top} = \frac{4 \cdot P_g}{\omega \cdot V_{Link}^2}$	$C_{bot} = C_{top} = \frac{4 \cdot P_g}{\omega \cdot V_{Link}^2}$	$C_{bot} < C_{top}$	$C_{bot} \gg C_{top}$
					$C_{top} - C_{bot} = \frac{2 \cdot P_g}{\omega \cdot V_{Link}^2}$	$C_{bot} - C_{top} = \frac{2 \cdot P_g}{\omega \cdot V_{Link}^2}$
Open loop PR, SVPWM	Close loop PR	Close loop	Open loop [16] Close loop [18] PR-50Hz	Close loop Multi - PR DC-offset	Close loop Multi-PR	Close loop Proposed multi- unbalanced αβ/dq DC-offset

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

of the proposed HSC-PD with other counterparts is shown in Table I.

IV. Experimental Results

The experimental results of the 7.4kW interleaved totem-pole PFC with HSC-PD in Fig.14 are demonstrated in this section to verify the above analysis and System design parameters summarized in Table II. The grid current, grid voltage, dc-link voltage, and decoupling inductor current at full load condition are shown in Fig. 15(a). Grid current and voltage are controlled in-phase with measured power factor of 0.9986. THD of the grid current is low of 3.6% due to interleaved approach even though each inductor current has a significantly higher THD of 5%, as shown in Fig. 15(b).

1) Decoupling capacitor and inductor performance

The decoupling capacitor voltages measured at half and full loads are shown in Fig. 16. The voltages across decoupling capacitor are fluctuating with the same frequency but 180° phase-shift. Dc-offset votlages of the bottom and top capacitors are set to 25% and 75% of the V_{Link}, resepctively, resulting in m = 0.25. DC-link voltage is flat with low ripple of 12.6V at full load meaning that the PD circuit functions well with low

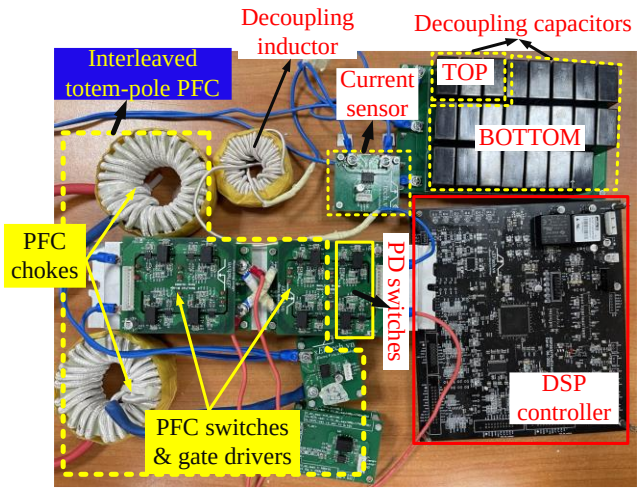
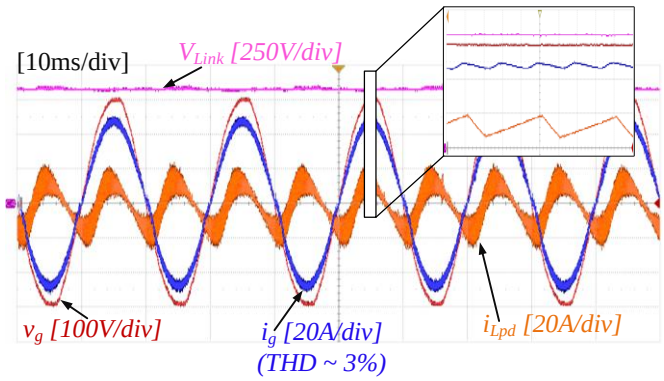


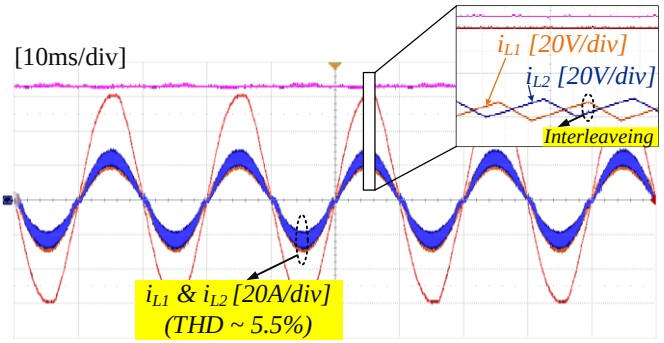
Fig. 14. Experimental prototype of the 7.4kW interleaved boost PFC and HSC-PD.

Table II
Designed Parameters of The Proposed Converter

Items	Values
Power rating	7.4 kW
Dc-link voltage, V_{Link}	820 V_{dc}
Grid voltage, v_g	220 V_{rms}
Switching frequency of PFC circuit	50 kHz
Switching frequency of PD circuit	50 kHz
Input inductors, L_1, L_2	500 µH
Decoupling inductor, L_{pd}	250 µH
Decoupling bottom capacitor, C_b	100 µF
Decoupling top capacitor, C_t	15 µF

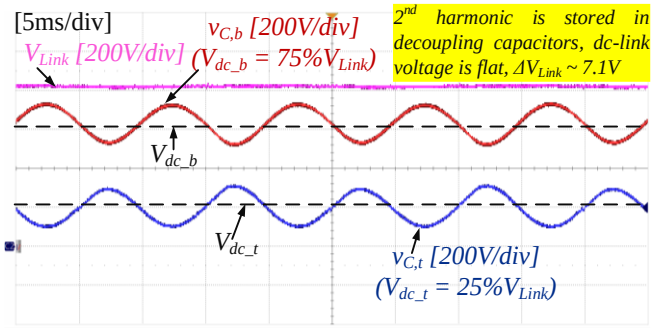


(a)

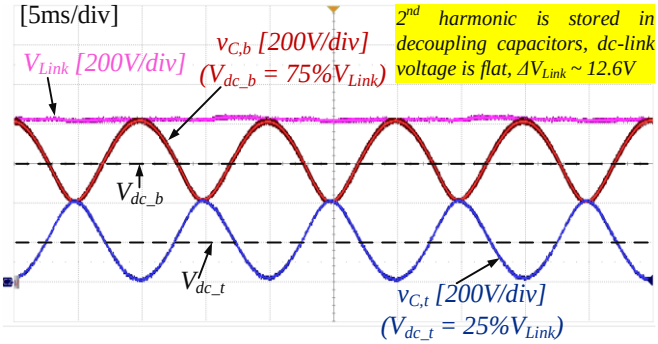


(b)

Fig. 15. Experimental results of the PFC circuit: (a) Grid current with low THD and high PFC is achieved at rated dc-link voltage, and the double-line frequency appears at the decoupling inductor current; (b) interleaved approach is achieved at the input currents.



(a)



(b)

Fig. 16. Decoupling capacitor voltage: (a) 50% of load; (b) full load

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

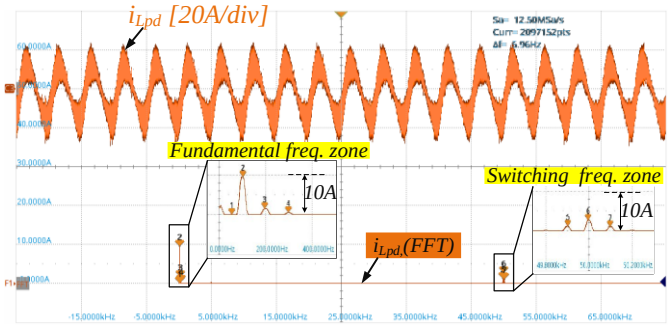


Fig. 17. Decoupling inductor current and its FFT analysis.

dc-link capacitance, and the 2nd harmonic is almost cancelled in both cases 50% and 100% of load. The ac components of the decoupling capacitors voltage at 50% of load has lower magnetude than that of the 100% of load.

The FFT analysis of the decoupling inductor current is shown in Fig. 17. There are two different harmonics zones; the fundermantal frequency zone around 50Hz, and the other harmonic zone is at around swtiching frequency of 50kHz. It can be seen from Fig. 17 that, the amplitude of 2nd harmonic is dominant over other components, verified the theoretical analysis, as shown in Fig. 4.

A step load changes from half to full load is shown in Fig. 18. In general, the transient of HSC-PD capacitors and inductor is smooth. The dc-link voltage drops within a short period of time at the transient because it is regulated by conventional dual-loop controller of the totem-pole PFC circuit without any effort to support the fast transient in addition to low equivalent dc-link capacitance and non-ideal grid in our facility. However, it worth to mention that fast dynamic response is not typically a critical requirement in OBC application since the battery voltage or battery current changes very gradually over an extended charging period. Therefore, the design in this paper does not specifically target this performance aspect.

The method based on conventional PR-controller in [15] and the proposed methods are both implemented on a DSP TMS320F28377D microprocessor, and the experimental comparison of the computation times for the two methods is shown in Fig.19. The results clearly highlight the computational efficiency of the proposed method, which outperforms the conventional approach with a faster calculation time of 389ns.

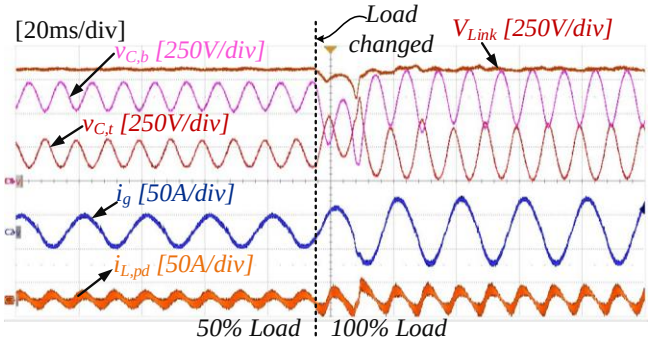


Fig. 18. Transient response of the step load changed from 50% to 100% of load.

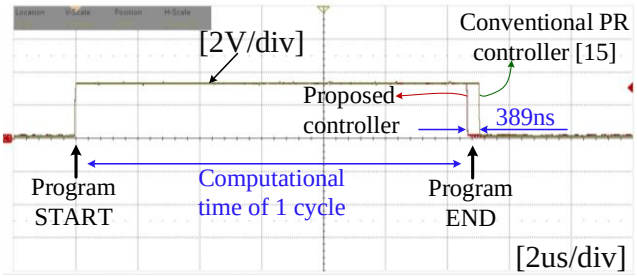


Fig. 19. Computation times for the practical implementation of both the conventional and proposed methods on a DSP TMS320F28377D.

2) Comparisons between the proposed HSC-PD and conventional E-cap method

The measured efficiency of the conventional totem pole PFC with E-cap and the proposed abroad with HSC-PD is shown in Fig. 20, full-load efficiency drops about 0.5% with PD due to additional losses of the HSC-PD circuit. The dropped efficiency is more significant under light load condition because the additional switching loss and core loss of the HSC-PD circuit are more dominant other conduction losses. It is however worth to mention that, in OBC applications the

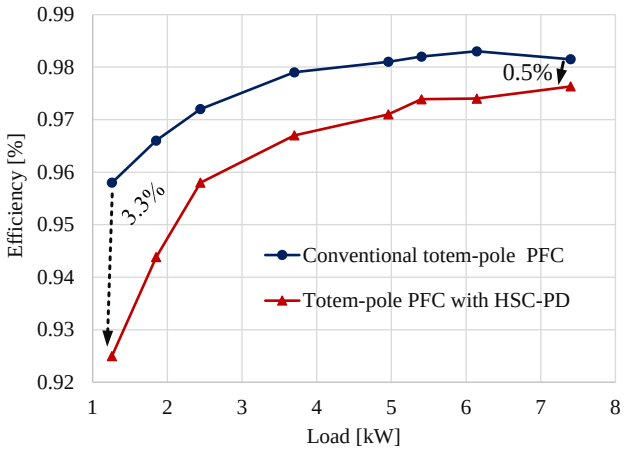


Fig. 20. Measured efficiency of the totem-pole PFC with and without PD circuit.

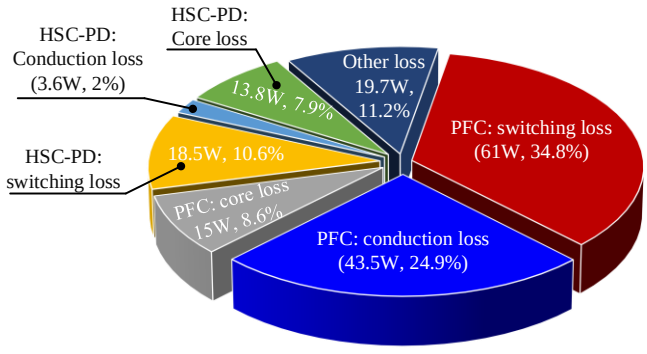


Fig. 21. Loss distribution of the Interleaved Totem-pole PFC with HSC-PD at V_g = 220V, V_{Link} = 820V, P_o = 7.4kW, f_s = 50kHz.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

TABLE III
SELECTED DEVICES FOR HSC-PD AND CONVENTIONAL E-CAP METHOD

Component	HSC-PD	E-cap
Capacitor	23x BLH505K901A062	148x 450BXW68MEFR18X25
Switch	2x NTHL040N120SC1	
Gate driver IC Power supply	2x UCC5320SCD 2x PDS2-S12-S24-M-TR	
Core	1x CH400060	

converter operates under heavy load most of the time during one charging cycle.

The loss distribution of the converter at the full load is shown in Fig. 21. Due to hard-switching condition, switching loss of main switches of the interleaved totem-pole PFC is the most dominant loss. The HSC-PD contributes 20.5% to the total loss of the converter at full-load condition. Within the HSC-PD, the switching loss and inductor core loss dominant over the conduction loss due to low current rating of the decoupling circuit.

Furthermore, with the selected components are presented in Table III, total cost and volume of the proposed HSC-PD are 1.9 and 2.2 times lower than that of the conventional E-cap method, respectively, as demonstrated in Fig. 22. Therefore, while acknowledging a marginal reduction in efficiency, the substantial cost and volume savings, complemented by improved reliability, underscore the advantageous trade-off of our proposed HSC-PD approach. This underlines the HSC-PD's potential as a viable alternative to traditional methods, paving the way for more compact and cost-effective OBC.

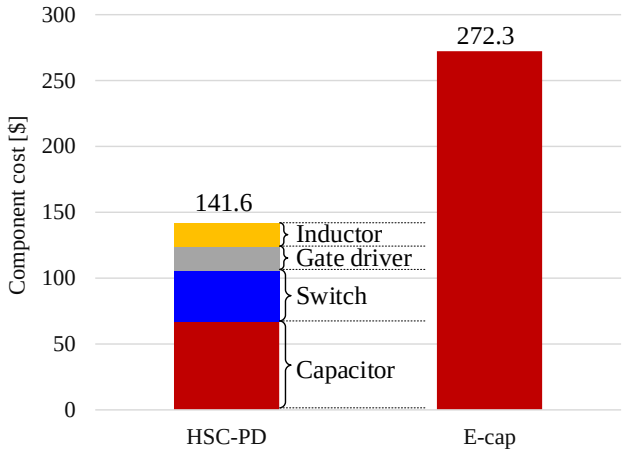


Fig. 22. Cost and volume comparison between the proposed HSC-PD and E-cap: (a) cost; (b) volume. (Cost is based on digikey.com accessed Jan. 2023)

3) Inrush current and precharge resistor

In OBC application, precharge circuit, which typically consits of one relay and one precharge resistor, is usually used at the input side to limit the inrush current when the OBC is connected to the grid, as seen in Fig. 23(a). The equivalent circuit of the converter when OBC is connected to the grid without and with precharge resistor is shown in Fig. 23(b) and Fig. 23(c), respectively. Assumming grid voltage is fixed, the rating of precharge resistor is mostly based on dc-link capacitance so that the higher dc-link capacitance the higher power needed of the precharge resistor. Therefore, in case of

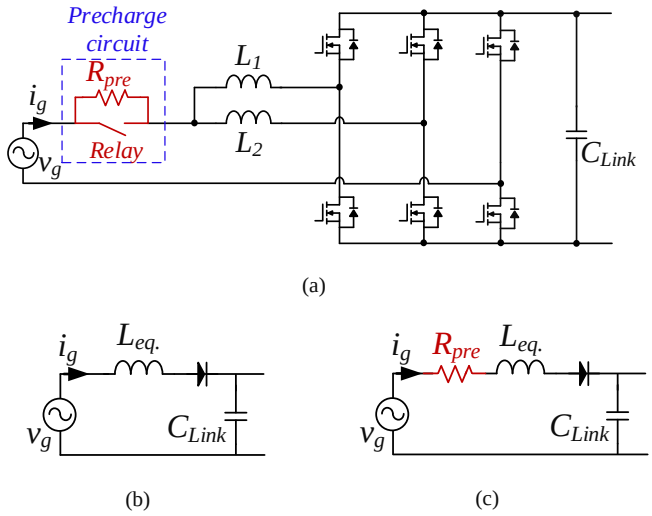


Fig. 23. Precharge circuit and equivalent circuit of AC/DC converter during start-up: (a) interleaved totem-pole PFC with precharge circuit; (b) equivalent PFC circuit during start-up without precharge resistor; (c) equivalent PFC circuit during start-up with precharge resistor.

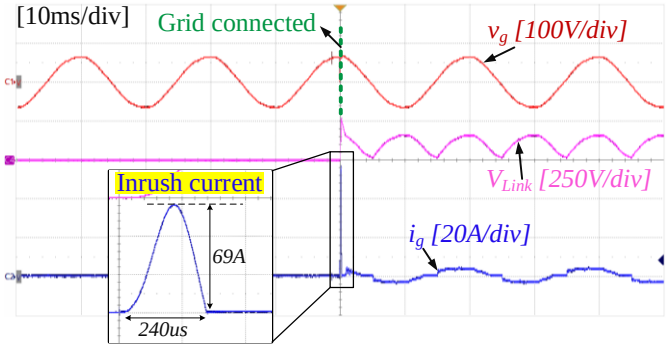


Fig. 24. Experimental waveforms of the inrush current when connecting the HSC-PD based PFC circuit with low dc-link capacitance to the grid without precharge resistor.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

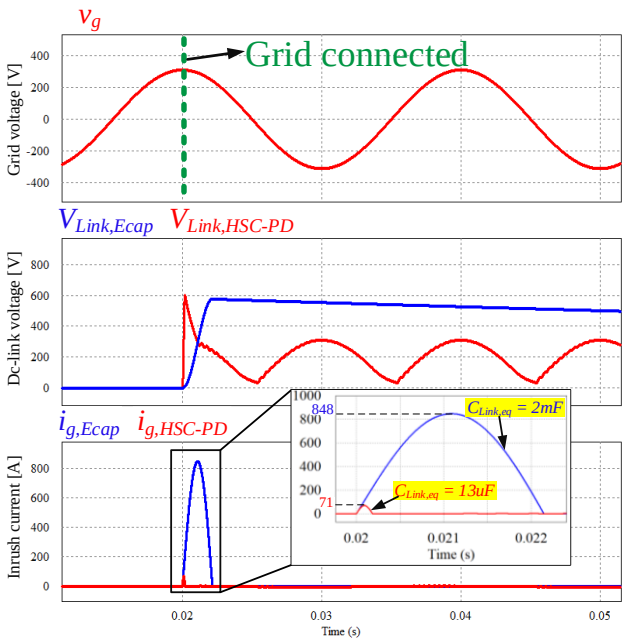


Fig. 25. Inrush currents comparison between conventional PFC with large E-cap and proposed HSC-PD based PFC with low capacitance F-cap.

using PD circuit, the precharge resistor can be significantly reduced compared to that of the conventional rectifier with large E-cap, reducing volume and cost.

Fig. 24 shows the input inrush current of the totempole PFC using HSC-PD circuit. The inrush current is within the acceptable magnetute and short interval due to small dc-link capacitance. The experiment with conventional totempole using large dc-link capaciance with E-cap could not be performed without prechage circuit due to safety issue. However, the simulation results of the inrush current in two cases are shown in Fig. 25. It can be seen that the conventional rectifier with large dc-link capacitor has significantly larger inrush current compared to that of the proposed converter.

VI. Conclusions

In this paper, a new method for designing and controlling a Half-bridge Split Capacitor based Power Decoupling circuit (HSC-PD) is presented with the dominant of bottom capacitor over the top one. The proposed design approach is demonstrated to have the following addvantage over the previous works:

1) Smaller capacitance value of each decoupling capacitors
2) Smaller equivalent dc-link capcitance
3) Reducing the required ac voltage rating of the film capacitor, which is critical in high voltage applications such as 800V OBC due to the low availability and costly of the high ac voltage rating film capacitor.
4) Smaller decoupling inductor current

Moreover, the small signal model of the HSC-PD is presented, and the modified the feedback control of the HSC-PD is proposed based on the co-ordinate transformation αβ/dq, which reduces the memory utilization and calculation effort of the DSP microcontroller compared to the conventional PR implementation. In addition, the effect of dc-link capacitance to precharge resistor power rating is investigated, showing the additional advantage of proposed method. Finally, experimental results of the 7.4kW 820V interleaved totem-pole PFC with HSC-PD are given to demonstrate the superiorities of the proposed method.

While our proposed HSC-PD scheme presents several significant advantages, it does come with its own unique set of challenges. These include a marginal decrease in efficiency due to the incorporation of additional circuit elements, increased complexity in PCB and control design, potential limitations for applications requiring prolonged hold-up times due to low dc-link capacitance value, and meticulous selection requirements for capacitors, especially considering their current rating due to low capacitance of the top capacitor. These issues offer ample opportunities for further refinement and optimization in future research.

References

[1] D. G. Barroso, Y. Yang, F. A. Machado and A. Emadi, "Electrified Automotive Propulsion Systems: State-of-the-Art Review," in *IEEE Transactions on Transportation Electrification*, vol. 8, no. 2, pp. 2898-2914, June 2022, doi: 10.1109/TTE.2021.3131917.

[2] Y. Park, S. Chakraborty and A. Khaligh, "DAB Converter for EV Onboard Chargers Using Bare-Die SiC MOSFETs and Leakage-Integrated Planar Transformer," in *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 209-224, March 2022, doi: 10.1109/TTE.2021.3121172.

[3] M. Yilmaz and P. T. Krein, "Review of Battery Charger Topologies, Charging Power Levels, and Infrastructure for Plug-In Electric and Hybrid Vehicles," in *IEEE Transactions on Power Electronics*, vol. 28, no. 5, pp. 2151-2169, May 2013, doi: 10.1109/TPEL.2012.2212917.

[4] Global EV Outlook 2022 [Online]. Available: *https://iea.blob.core.windows.net/assets/ad8fb04c-4f75-42fc-973a-6e54c8a4449a/GlobalElectricVehicleOutlook2022.pdf* (Accessed: 02-March-2023).

[5] A. Khaligh and M. D'Antonio, "Global Trends in High-Power On-Board Chargers for Electric Vehicles," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3306-3324, April 2019, doi: 10.1109/TVT.2019.2897050.

[6] C. -Y. Oh, D. -H. Kim, D. -G. Woo, W. -Y. Sung, Y. -S. Kim and B. -K. Lee, "A High-Efficient Nonisolated Single-Stage On-Board Battery Charger for Electric Vehicles," in *IEEE Transactions on Power Electronics*, vol. 28, no. 12, pp. 5746-5757, Dec. 2013, doi: 10.1109/TPEL.2013.2252200.

[7] B. Li, F. C. Lee, Q. Li and Z. Liu, "Bi-directional on-board charger architecture and control for achieving ultra-high efficiency with wide battery voltage range," 2017 *IEEE Applied Power Electronics Conference and Exposition (APEC)*, Tampa, FL, USA, 2017, pp. 3688-3694, doi: 10.1109/APEC.2017.7931228.

[8] Z. Zhang, C. Liu, M. Wang, Y. Si, Y. Liu and Q. Lei, "*High-Efficiency High-Power-Density CLLC Resonant Converter With Low-Stray-Capacitance and Well-Heat-Dissipated Planar Transformer for EV On-Board Charger*," in IEEE Transactions on Power Electronics, vol. 35, no. 10, pp. 10831-10851, Oct. 2020, doi: 10.1109/TPEL.2020.2980313.

[9] J. Lu, A. Mallik and A. Khaligh, "Dynamic Strategy for Efficiency Estimation in a CCM-Operated Front-End PFC Converter for Electric Vehicle Onboard Charger," in *IEEE Transactions on Transportation Electrification*, vol. 3, no. 3, pp. 545-553, Sept. 2017, doi: 10.1109/TTE.2017.2731622.

[10] F. Musavi, M. Edington, W. Eberle and W. G. Dunford, "Evaluation and Efficiency Comparison of Front End AC-DC Plug-in Hybrid Charger

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

Topologies," in *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 413-421, March 2012, doi: 10.1109/TSG.2011.2166413.

[11] U.S. Department of Energy (2017), "Electrical and Electronics Technical Team Roadmap," [Online]. Available: https://www.energy.gov/sites/prod/files/2017/11/f39/EETT%20Roadmap%2010-27-17.pdf (Accessed: 10-March-2023).

[12] H. Kim, J. Park, S. Kim, R. M. Hakim, H. Belkamel and S. Choi, "A Single-Stage Electrolytic Capacitor-Less EV Charger With Single- and Three-Phase Compatibility," in *IEEE Transactions on Power Electronics*, vol. 37, no. 6, pp. 6780-6791, June 2022, doi: 10.1109/TPEL.2021.3127010.

[13] C. Jung, "Power Up with 800-V Systems: The benefits of upgrading voltage power for battery-electric passenger vehicles," in *IEEE Electrification Magazine*, vol. 5, no. 1, pp. 53-58, March 2017, doi: 10.1109/MELE.2016.2644560.

[14] R. Wang et al., "A High Power Density Single-Phase PWM Rectifier With Active Ripple Energy Storage," in *IEEE Transactions on Power Electronics*, vol. 26, no. 5, pp. 1430-1443, May 2011, doi: 10.1109/TPEL.2010.2090670.

[15] K. -P. Huang, Y. Wang and R. -J. Wai, "Design of Power Decoupling Strategy for Single-Phase Grid-Connected Inverter Under Nonideal Power Grid," in *IEEE Transactions on Power Electronics*, vol. 34, no. 3, pp. 2938-2955, March 2019, doi: 10.1109/TPEL.2018.2845466.

[16] Y. Tang, F. Blaabjerg, P. C. Loh, C. Jin and P. Wang, "Decoupling of Fluctuating Power in Single-Phase Systems Through a Symmetrical Half-Bridge Circuit," in *IEEE Transactions on Power Electronics*, vol. 30, no. 4, pp. 1855-1865, April 2015, doi: 10.1109/TPEL.2014.2327134.

[17] W. Yao, P. C. Loh, Y. Tang, X. Wang, X. Zhang and F. Blaabjerg, "A Robust DC-Split-Capacitor Power Decoupling Scheme for Single-Phase Converter," in *IEEE Transactions on Power Electronics*, vol. 32, no. 11, pp. 8419-8433, Nov. 2017, doi: 10.1109/TPEL.2016.2645800.

[18] Y. Tang, Z. Qin, F. Blaabjerg and P. C. Loh, "A Dual Voltage Control Strategy for Single-Phase PWM Converters With Power Decoupling Function," in *IEEE Transactions on Power Electronics*, vol. 30, no. 12, pp. 7060-7071, Dec. 2015, doi: 10.1109/TPEL.2014.2385032.

[19] R. Chen, Y. Liu and F. Z. Peng, "DC Capacitor-Less Inverter for Single-Phase Power Conversion With Minimum Voltage and Current Stress," in *IEEE Transactions on Power Electronics*, vol. 30, no. 10, pp. 5499-5507, Oct. 2015, doi: 10.1109/TPEL.2014.2375271.

[20] W. Qi, H. Wang, X. Tan, G. Wang and K. D. T. Ngo, "A novel active power decoupling single-phase PWM rectifier topology," 2014 *IEEE Applied Power Electronics Conference and Exposition* - APEC 2014, 2014, pp. 89-95, doi: 10.1109/APEC.2014.6803293.

[21] H. Li, K. Zhang, H. Zhao, S. Fan and J. Xiong, "Active Power Decoupling for High-Power Single-Phase PWM Rectifiers," in *IEEE Transactions on Power Electronics*, vol. 28, no. 3, pp. 1308-1319, March 2013, doi: 10.1109/TPEL.2012.2208764.

[22] H. Zhao, Y. Shen, W. Ying, S. S. Ghosh, M. R. Ahmed and T. Long, "A Single- and Three-Phase Grid Compatible Converter for Electric Vehicle On-Board Chargers," in *IEEE Transactions on Power Electronics*, vol. 35, no. 7, pp. 7545-7562, July 2020, doi: 10.1109/TPEL.2019.2956653.

[23] H. Sun, H. Wang and W. Qi, "Automatic Power Decoupling Controller of Dependent Power Decoupling Circuit for Enhanced Transient Performance," in *IEEE Transactions on Industrial Electronics*, vol. 66, no. 3, pp. 1820-1831, March 2019, doi: 10.1109/TIE.2018.2838100.

[24] Y. Wang and R. -J. Wai, "Adaptive Power Decoupling Strategy for Single-Phase Grid-Connected Converter," in *IEEE Transactions on Industry Applications*, vol. 55, no. 4, pp. 4275-4285, July-Aug. 2019, doi: 10.1109/TIA.2019.2908945.

[25] Y. Sun, Y. Liu, M. Su, W. Xiong and J. Yang, "Review of Active Power Decoupling Topologies in Single-Phase Systems," in *IEEE Transactions on Power Electronics*, vol. 31, no. 7, pp. 4778-4794, July 2016, doi: 10.1109/TPEL.2015.2477882.

[26] R76, Double Metallized Polypropylene Film, Radial, DC and Pulse Applications (Automotive Grade) [Online]. Available: https://content.kemet.com/datasheets/KEM_F3034_R76.pdf (Accessed: 08-March-2023).

[27] Type BLH, DC Link Capacitors for Harsh Environments [Online]. Available: https://www.mouser.com/datasheet/2/88/BLH-3081877.pdf (Accessed: 08-March-2023).

[28] F. Hans, W. Schumacher, S. -F. Chou and X. Wang, "Design of Multifrequency Proportional–Resonant Current Controllers for Voltage-Source Converters," in *IEEE Transactions on Power Electronics*, vol. 35, no. 12, pp. 13573-13589, Dec. 2020, doi: 10.1109/TPEL.2020.2993163.

[29] C. Zou, B. Liu, S. Duan and R. Li, "Stationary Frame Equivalent Model of Proportional-Integral Controller in dq Synchronous Frame," in *IEEE Transactions on Power Electronics*, vol. 29, no. 9, pp. 4461-4465, Sept. 2014, doi: 10.1109/TPEL.2013.2296789.

[30] F. J. Rodriguez, E. Bueno, M. Aredes, L. G. B. Rolim, F. A. S. Neves and M. C. Cavalcanti, "Discrete-time implementation of second order generalized integrators for grid converters," 2008 *34th Annual Conference of IEEE Industrial Electronics*, Orlando, FL, USA, 2008, pp. 176-181, doi: 10.1109/IECON.2008.4757948.

[31] A. G. Yepes, F. D. Freijedo, J. Doval-Gandoy, Ó. López, J. Malvar and P. Fernandez-Comesaña, "Effects of Discretization Methods on the Performance of Resonant Controllers," in *IEEE Transactions on Power Electronics*, vol. 25, no. 7, pp. 1692-1712, July 2010, doi: 10.1109/TPEL.2010.2041256.

[32] D. Dong, T. Thacker, R. Burgos, F. Wang and D. Boroyevich, "On Zero Steady-State Error Voltage Control of Single-Phase PWM Inverters With Different Load Types," in *IEEE Transactions on Power Electronics*, vol. 26, no. 11, pp. 3285-3297, Nov. 2011, doi: 10.1109/TPEL.2011.2157361.

Tuyen Dinh Nguyen (*Senior Member, IEEE*) was born in Binh Dinh, Vietnam, in 1982. He received the B.S. degree in electrical engineering from the Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2004, and the Ph.D. degree from the University of Ulsan, Ulsan, South Korea, in 2012. He is currently an Associate Professor at the Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology, Vietnam National University. His research interests include power electronics converter, power electronics system integration, its industrial applications such as renewable energy and electric vehicle.

Long H. B. Nguyen received the B.S. degrees in Electrical Engineering in 2022 from HCMC University of Technology (Bach Khoa University), Ho Chi Minh City, Vietnam, where He is currently working toward the M.S. degree in Power Electronics. Since 2022, He has been a Power Electronics Engineer with xEVtech Co. Ltd., Ba Ria – Vung Tau, Vietnam. His research interests include PFC converter topologies, Power Decoupling algorithm, EV chargers.

Dong Tran received the B.S. degrees in Electrical Engineering in 2022 from HCMC University of Technology (Bach Khoa University), Ho Chi Minh City, Vietnam. Since 2022, He has been a Power Electronics Engineer with xEVtech Co. Ltd., Ba Ria – Vung Tau, Vietnam. His research interests include PFC converter topologies and Power Decoupling control algorithm.

Ravi Nath Tripathi (*Senior Member, IEEE*) received a B.Tech degree in Electrical and Electronics Engineering from Graphic Era Institute of Technology (presently Graphic Era Deemed to be University), Dehradun, India, in 2011 and M.tech degree from Delhi Technological University, Delhi, India in 2013. He received his PhD degree from Kyushu Institute of Technology, Fukuoka, Japan in 2017 and after that worked as a post-doctoral researcher on the NEDO project from 2017-2020 at Next Generation Power Electronics Research Center, Kyushu Institute of Technology. He joined the Kyoto University of Advanced Science, Kyoto, Japan in 2020 as an assistant professor and currently working as an assistant professor at the Kyushu Institute of Technology. His research

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Transportation Electrification. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TTE.2023.3299988

interest is gate driving and system control for power semiconductor devices, power converter system development and control, renewable energy systems and virtual prototyping.

Hai N. Tran (M'21) received the B.S. and M.S. degrees in Electrical Engineering from HCMC University of Technology (Bach Khoa University), Ho Chi Minh city, Vietnam, in 2014 and 2016, respectively, the Ph.D. degree from Seoul National University of Science and Technology (SeoulTech), Seoul, Korea, in 2021. He was a Postdoctoral Researcher with the Nagamori Institute of Actuators, Kyoto University of Advanced Science, Kyoto, Japan in 2021, and a Postdoctoral Fellow with University of South Carolina, South Carolina, USA in 2022. He is a founder of xEVtech Co. Ltd., Ba Ria – Vung Tau, Vietnam, where is currently a CTO. His research interests include active snubber circuits, soft switching converter topologies, medium voltage, EV chargers, and magnetic design.
Dr. Tran received the 2022 IEEE Transactions on Power Electronics (TPEL) Prize Paper Award.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 10,2023 at 06:46:25 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 July 2023, accepted 7 August 2023, date of publication 22 August 2023, date of current version 30 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3307388

RESEARCH ARTICLE

Using Dielectric Rods to Reconfigure Combline Filters in Center Frequency and Bandwidth

ABHISHEK SHARMA, (Student Member, IEEE), SANTIAGO COGOLLOS, (Member, IEEE), VICENTE E. BORIA, (Fellow, IEEE), AND MARCO GUGLIELMI, (Life Fellow, IEEE)
Instituto de Telecomunicaciones y Aplicaciones Multimedia (ITEAM), Universitat Politècnica de València, 46022 Valencia, Spain

Corresponding author: Abhishek Sharma (asharma@iteam.upv.es)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program through the Marie Skłodowska-Curie Innovative Training Networks (ITN) (TESLA) under Grant 811232; and in part by the Minsiterio de Ciencia e Innovación (MICIN), Spanish Government, through the Research and Development Project under Grant PID2019-103982RB-C41 (funded by MICIN/AEI/10.13039/501100011033).

ABSTRACT This article describes the use of dielectric rods to reconfigure combline filters both with respect to center frequency and bandwidth. In this context, a new filter topology is developed in order to maximize the achievable reconfiguration range. The design of a prototype is discussed in detail, clearly demonstrating the feasibility of center frequency and bandwidth reconfiguration, with the same filter structure, while maintaining an equiripple return loss. In addition to theory, a combline filter prototype is also fabricated, tuned and measured in several configurations showing excellent agreement with simulations, thereby fully validating both the basic concept and the design procedure.

INDEX TERMS Band-pass filter, combline filters, dielectric tuners, reconfigurability, tunable filters.

I. INTRODUCTION

The current evolution trend in communication equipment requirements, for both ground and space applications, is towards an ever increasing use of bandwidth. However, the total available bandwidth is limited, so that the frequency spectrum that is allocated to each application is now fully regulated at world wide level [1]. To meet this challenge, modern communication systems are currently being developed with the ability of providing multi-channel communication systems, where each channel uses different center frequency and bandwidth [2].

In this context therefore, the possibility of developing microwave components, and filters in particular, that are agile in frequency has recently become a very important area of research [3], [4], [5]. Filters (and multiplexers) have always been considered key components for the front-end of communication equipment for both ground and space applications [6]. As a consequence, a very significant research effort has been devoted in the last decade to the development of filters and multiplexers that can be reconfigured both in terms of center frequency and bandwidth [7], [8].

The simplest approach to change the center frequency of a filter structure based on coupled cavities, is to change the size of the cavities. A number of different approaches that explore this possibility have indeed been reported in the technical literature. For instance, a compressible bellow or a conductive elastomer lid on the top of two dual-mode cavities has been demonstrated in [9]. A different concept has also been proposed in [10] discussing a contact-less corrugated plunger. Furthermore, one more approach is presented in [11], where the cavity size is changed by activating or deactivating an array of conductive lines located on the cavity wall. An additional approach has also been introduced in [12], for the reconfiguration of both center frequency and bandwidth by moving, at the same time, both end walls of a cylindrical cavity. Two more contributions can also be found in the technical literature, where the filter tunability is obtained by rotating a specially designed piece of dielectric material that is located inside the resonant cavities of the filter [13], [14].

Furthermore, in [15] a copper coated polyflon sheet is used to cover the top of a metallic cavity, and a piezoelectric

The associate editor coordinating the review of this manuscript and approving it for publication was Ye Zhou.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

actuator is used to modify the cavity dimensions. Another approach is described in [16], where tuning disks are actuated using thermal MEMS. Several papers have also discussed capacitive coupling between resonator and tuning discs or posts. In [17] and [18], for instance, a floating capacitance between the disk and the post is modified using a MEMS-based capacitance or a varactor. One more approach is described in [19] where GaAs and MEMS capacitors are proposed. In [20], another tuning solution is adopted, namely, a varactor diode connected to the open end of a coaxial post resonator.

Another very common approach to tune filters based on waveguide cavities is to use metallic tuning elements or screws. This approach has been demonstrated with directly-coupled rectangular waveguide filters [21], and dual-mode filters [22]. Classical combline filter topologies can also be tuned with metallic screws, as it is shown in [23] and [24]. More recently, metal screws have also been used within re-entrant coaxial resonators, for tuning the frequency response while keeping the bandwidth constant [25]. However, the use of metallic tuning elements (or screws) for remotely reconfigurable filter implementations, can introduce a number of complications, namely:

- The tuning element or screw must be securely fastened to the filter body in order to operate correctly.
- The penetration of the tuning screw must be controlled very accurately to ensure the correct filter performance.
- The presence of metallic tuning elements may have a negative effect on the high power and passive intermodulation (PIM) performance of the filter.

Alternative solutions, based on the use of a single rotating mechanism driving all metal tuning elements, have been recently developed for E-plane filters [26], and coaxial filters [27]. These solutions have a simplified mechanical assembly, and also allow to keep a constant bandwidth in the achievable frequency tuning ranges. This is achieved using a clever design for the coupling elements of the filter. However, these new topologies can only produce filters with fixed bandwidth (approximately 2%), with a return loss of 15 dB.

In addition to metallic tuners, however, dielectric tuners have also been used more recently to tune microwave filters. In [28] and [29], their potential use for tuning standard rectangular waveguide filters is validated through full-wave simulations and measurements. A very recent work on the practical demonstration of dielectric tuners in inductive waveguide filters is discussed in [30]. It is important to note, at this point, that the results discussed in [30] are only applicable to rectangular waveguide filters based on *distributed* resonators. The present work, in contrast, intends to demonstrates the use of dielectric tuners in combline filters, where the resonators can essentially be considered as *lumped-element* resonators. In the context of combline filters, dielectric tuners were first used within resonators, in [31] and [32], for implementing tunable filters in terms of center frequency only (keeping constant the filter bandwidth). Additional initial results for the same tunable responses (only in terms of center frequency) have also been discussed in [33], where dielectric tuners are used in coupling windows to preserve the same filter bandwidth.

In this context, therefore, the objective of this paper is to significantly extend the initial results presented in [33]. The additional novel material discussed in this paper is as follows:

- In addition to Alumina, we also study the use of Zirconia, and we compare the key performance of the two dielectric materials.
- A new filter structure is then designed to take advantage of the superior performance of Zirconia.
- Furthermore, we show that both simulation and measurements indicate that Zirconia has a good quality factor, that is comparable to other established topologies using either metallic or dielectric tuners.
- In addition to changing the center frequency, we show how the same structure can also be used to change the bandwidth of the filter. In particular, we show that the combline filter can be reconfigured to maintain a constant absolute bandwidth throughout the frequency tuning range. Alternatively, the bandwidth of the filter can be increased up to twice the initial value.
- We show that the relation between the variation of the penetration of the tuning elements and the filter center frequency remains approximately linear. As a consequence, the prediction of the filter center frequency is greatly simplified.
- We discuss how the use of dielectric tuners eliminates the need for keeping a very good electrical contact with the body of the filter, thereby eliminating a common source of insertion loss and PIM.
- In addition to simulations, we also show the measured results of a prototype built using Zirconia, and we demonstrate successfully how the same filter structure can be reconfigured both in terms of center frequency and bandwidth.
- Furthermore, we show how the out of band performance of tunable filter can be improved with the integration of a simple lowpass filter. In this context, we also discuss the achievable improvements and limitations of this approach.
- In addition, we also study the high power behavior of our tunable filter, and discuss the possible trade-offs between power handling and tuning range.
- Finally, we discuss how the structure we propose opens the possibility of using low-accuracy, computer-controlled actuators to implement remotely controlled reconfiguration of the filter.

This article is organized as follows: In section II, the design of a frequency-tunable filter using dielectric rods made of Alumina is discussed in detail. We then discuss in section III the use of Zirconia as an alternative dielectric material, and derive a new set of initial filter dimensions. After that, we discuss in section IV the complete design of a filter that can be tuned both in terms of frequency and bandwidth.

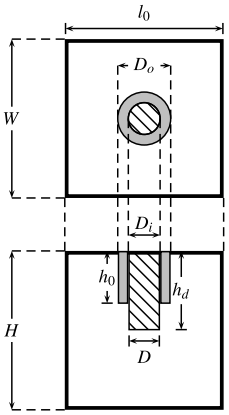


FIGURE 1. Cavity configuration where the dielectric tuner is inserted in the hollow post (top view and side view [33]).

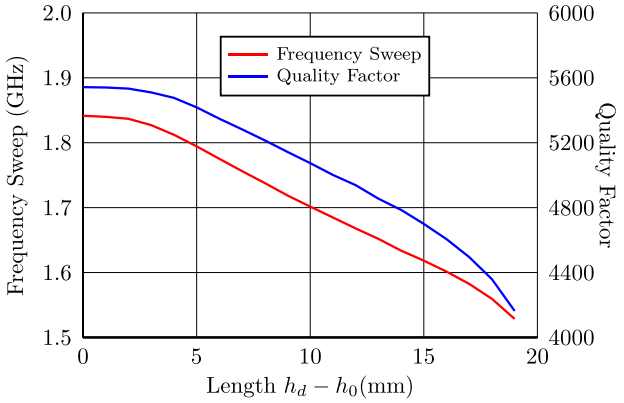


FIGURE 2. Frequency tuning range (left side scale) and Q_u factor (right side scale) provided by the structure in Fig. 1 as a function of the length of the dielectric rod exceeding the hollow post ($h_d - h_0$).

The details in the manufacturing process and the measured results are shown in sections V. In Section VI we discuss how the out-of-band response of the reconfigurable filter can be controlled. In Section VII we study the multipactor behavior of our filter. Section VIII describes the remote operation of the reconfigurable filter prototype. Finally, Section IX concludes the paper with a summary of the results achieved.

II. FREQUENCY-TUNABLE FILTER WITH ALUMINA DIELECTRIC TUNERS

In this section, we briefly recall, for the sake of completeness, the design of the frequency tunable filter based on Alumina tuners already discussed in [33], providing, at the same time, new information concerning the dimensions of the structure, and the evaluation of losses and Q-factors.

A. RESONATOR DESIGN CONSIDERATIONS

The resonators are the basic elements of any coupled cavity filter. It is common practice to initiate the design procedure by optimizing the structure of the resonator to operate in the desired resonant mode and frequency range. The selection of the adequate resonator topology is indeed a fundamental step because it has a strong effect on the unloaded quality factor (Q_u), filter size, cost, and power handling capability of the complete filter.

Different cavity configurations for the resonators have been studied in [33], and the most suitable design has been chosen (see Fig. 1). Once the basic shape of the resonator has been defined, metallic screws are normally placed in the vicinity of the open-ended post to create a variable capacitance to tune the resonance at the desired frequency.

There are, however, several issues associated with the use of metallic tuning screws. Namely, the metallic screws must be securely fastened to the body of the filter. As a consequence, remote tuning becomes problematic. Additionally, even if a remote tuning mechanism could be developed, the structure would be extremely sensitive to the position of the metallic tuning screw, so that a practical (low-cost) solution for an actuator-controlled remote tuning would be very difficult to implement. Finally, metallic tuning screws can also introduce significant losses in the filter structure if repeated tuning is required. In our basic resonator, we have therefore replaced the traditional metal screws with dielectric tuners.

Additionally, as we can see in Fig. 1, a hollow metallic post has been chosen instead a solid one [32]. This has been done because a hollow post provides a support for the movement of the dielectric tuning element. Another reason for choosing this configuration is the almost-linear variation of the frequency with the tuner penetration. Using this configuration, we can, in fact, obtain the desired starting value of the resonant frequency using the metallic post, and we can then have an extended tuning range inserting a dielectric rod inside the cavity using the hollow post as a support.

The dimensional parameters of the basic resonator in Fig. 1 are: $l_0 = H = 50\,\text{mm}$ and $W = 40\,\text{mm}$. The outer and inner diameters of the hollow post are $D_o = 16\,\text{mm}$ and $D_i = 12\,\text{mm}$, respectively. The height of the hollow post is $h_0 = 32.290\,\text{mm}$. h_d is the height of the dielectric tuner, the diameter of which is $D = 11.8\,\text{mm}$ (slightly less than D_i).

Fig. 2 shows the variation in center frequency with the tuner penetration. The linear pattern is clearly evident. Fig. 2 shows also the unloaded quality factor Q_u as a function of the dielectric post penetration. The values for the Q_u are quite high ranging from 4200 for deep penetrations up to near 5600 for penetrations not exceeding the hollow post h_0. It is interesting to note that the equivalent structure with a metallic tuner has a quality factor of 2758 at the lowest frequency compared with the Q value of 4200 using the dielectric tuner. Furthermore, we have verified that, with the resonator we have designed, we can indeed tune the resonance in the entire frequency range of interest.

B. DESIGN CONSIDERATIONS OF INPUT AND OUTPUT COUPLING

In the context of tunable filters, it is very crucial to maintain the correct input/output coupling throughout the frequency tuning range in order to maintain constant bandwidth and

return loss levels. The Q_{ext} value is computed from the desired filter center frequency (f_0) and bandwidth (BW) using the following classical expression [6]:

$$Q_{\text{ext}} = \frac{f_0}{M_{S1}^2 BW} \tag{1}$$

where M_{S1} is the coupling matrix element corresponding to the input coupling of the filter.

From equation (1), we can now deduce the following:

1) For lower f_0 and constant BW a higher input coupling is required. Therefore, lower frequencies are more challenging for tunable filters.
2) For wider BW a lower Q_{ext} (higher input coupling) is also required. Therefore wider BW are more challenging for tunable filters.

As initially discussed in [33], we propose to use the structure shown in Fig. 3, where the resonator post and the input coupling post are attached to opposite walls. This geometrical configuration allows for a stronger electric coupling, thus providing the required coupling strength. The input coupling post shares the same mechanism as the tunable resonator: a dielectric rod is inserted inside the hollow post. This arrangement does indeed provide a variable capacitive coupling that depends on the penetration of the dielectric rod.

The dimensions of the structure in Fig. 3 have been chosen to be consistent with the dimensions of the basic resonator, namely, $l_0 = H = 50\,\text{mm}$ and $W = 40\,\text{mm}$, where the height of the hollow resonator and the coupling structure are $h_0 = 26.172\,\text{mm}$ and $h_i = 26.82\,\text{mm}$, respectively. The input coaxial feed is located at $h_c = 19.140\,\text{mm}$. The value of h_i is optimized to obtain the input coupling value that is required at the highest frequency, when the dielectric rod is fully inserted in the metallic post (Fig. 3). The inner conductor of the coaxial input is connected to the first metallic post at a height $h_c = 19.140\,\text{mm}$ which is 7.68 mm shorter than h_i. This has been done in order to ensure the proper mechanical robustness of the input structure. The outer diameter of the input/output metallic post is set at $d_0 = 8\,\text{mm}$. The inner diameter is $d_i = 6\,\text{mm}$ and the rod inserted inside has a diameter of 5.95 mm. A gap s equal to 2.0 mm has been left between the input post and the resonator post. The lowest coupling value is obtained when the dielectric rod is totally confined inside the metallic hollow post having height $h_{\text{id1}} = 25\,\text{mm}$.

A key point to be emphasized at this stage is that the input coupling in our combline cavity loads significantly the first resonator. This leads to a reduction in the height of the resonator post (please note that $h_0 = 26.172\,\text{mm}$ in Fig. 3, whereas $h_0 = 32.290\,\text{mm}$ in Fig. 1). This, in turn, allows for an enhanced range in the frequency tuning of the input resonator.

Fig. 4 shows the experiment carried out using the structure in Fig. 3. From this result, we can now see that, using (1), the Q_{ext} value required to maintain the $RL = 25\,\text{dB}$ at 1.842 GHz is around 18.5 while at 1.542 GHz (for $BW = 75\,\text{MHz}$), the Q_{ext} value required is 15.5. Fig. 4 confirms that our structure can provide the required coupling values in the frequency range of interest. Moreover, the variation of Q_{ext} slightly degrades the quality factor of the cavity (from 4280 to 3922), as also shown in Fig. 4.

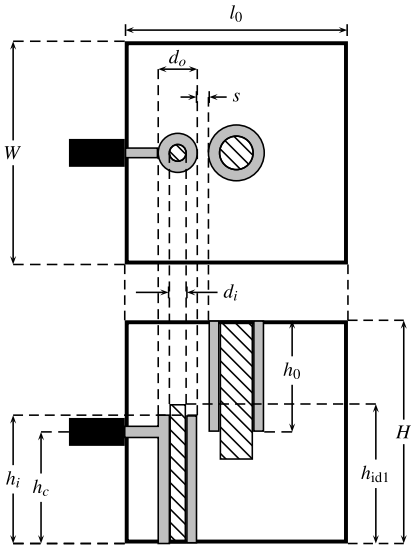


FIGURE 3. Input cavity configuration with the most significant geometrical parameters.

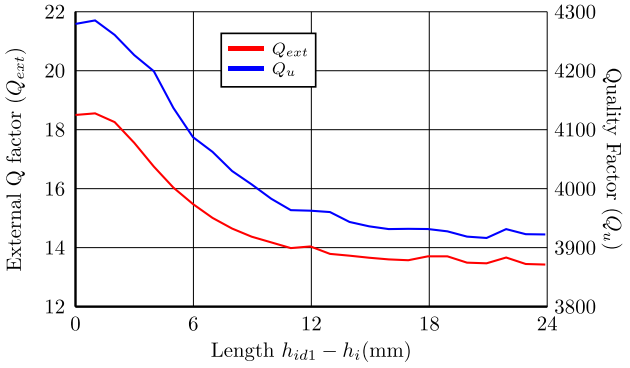


FIGURE 4. Q_{ext} (left side scale) and Q_u (right side scale) variation of the proposed solution. The x-axis shows the length of the dielectric rod exceeding the hollow post ($h_{\text{id1}} - h_i$).

An additional interesting point is that our simulations indicate that inserting the dielectric rod in the input coupling post lowers the resonant frequency of the cavity (it has been observed that it changes from 1.813 GHz for $h_{\text{id1}} - h_i = 0\,\text{mm}$ to 1.746 GHz for $h_{\text{id1}} - h_i = 24\,\text{mm}$). This is indeed a clear indication that confirms the capacitive nature of the input coupling.

C. DESIGN CONSIDERATIONS OF INTER-RESONATOR COUPLING

Generally speaking, a very common solution to implement the inter resonator coupling in combline filters is to use coupling windows (see for instance [34]). Metallic or dielectric tuning elements (or screws) are then located inside the coupling windows (2mm thick) to tune the coupling to the desired value (see the red inset in Fig. 5). However, our study

shows that using coupling windows and tuning elements is not adequate for the application under investigation. The reason is that the coupling value variations that can be achieved are insufficient for our application.

The solution that we propose in this paper consists of using a short hollow metallic post having an outer diameter of $d_0 = 8\,\text{mm}$, and an inner diameter of $d_i = 6\,\text{mm}$, with a dielectric tuner (of diameter 5.95 mm) inside, located on the wall opposite to the resonator post, in the middle between resonators (see the blue inset in Fig. 5). The rest of the dimensional variables (for the cavity resonator and dielectric tuner) are kept the same as the ones given in section II-A. The results in Fig. 5 show the coupling range obtained by full-wave simulation. The extraction of the coupling values has been carried out with two resonators weakly coupled to the input/output ports and strongly coupled to each other, as described in [6]. Another advantage of this solution is that the initial coupling value is then established with the height of the hollow metallic post $h_c = 11\,\text{mm}$. Another feature of this solution is that without a coupling window, the dielectric rod has more room to penetrate in the filter cavity, thus implementing a stronger coupling between resonators, as well as a wider range of coupling values. This is indeed confirmed by a comparison of the results shown in Fig. 5.

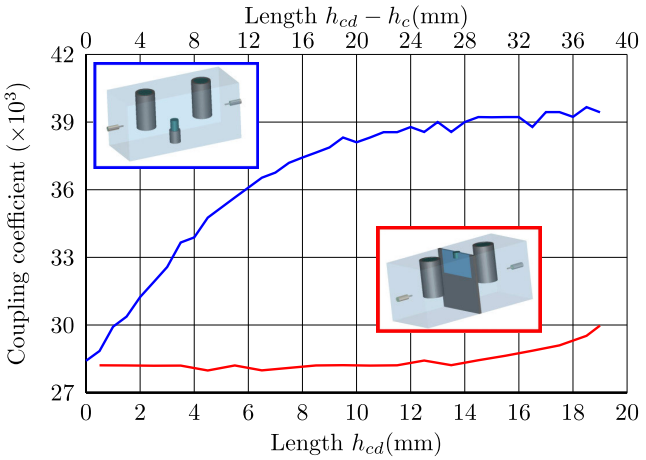


FIGURE 5. Inter-resonator coupling using a metallic post without a coupling window (upper side scale). The length of the hollow post holding the tuner is $h_c = 11$ mm. Length h_{cd} is the height of the dielectric tuner. Inter-resonator coupling using an iris with a dielectric tuning screw (lower side scale). h_{cd} is the height of the dielectric tuner. Blue line: response of coupling using a dielectric post. Red line: response of coupling using an iris.

D. COMPLETE FILTER GEOMETRY

The results that have been shown so far can effectively be used to establish the initial dimension of our filter structure.

The filter transfer function we want to implement is a fourth-order Chebyshev function. The desired specifications are as follows:

- Center frequency must be tunable from $f_{0,\min} = 1.542\,\text{GHz}$ up to $f_{0,\max} = 1.842\,\text{GHz}$.
- Bandwidth $BW = 75\,\text{MHz}$ constant throughout the whole range of frequencies.
- Return loss: $RL > 25\,\text{dB}$.

and must be achieved while maintaining a constant bandwidth and return loss performance. This means that the filter tuning range, as defined in [35], [36], and [37], for instance, is:

$$\frac{f_{0,\max} - f_{0,\min}}{f_C} \times 100 \approx 18\% \tag{2}$$

where f_C is given by:

$$\frac{f_{0,\max} + f_{0,\min}}{2} = 1.692\,GHz \tag{3}$$

and the relative bandwidth at f_C is $\approx$ 4.4%. Which means that the filter must be tuned over a frequency range that is approximately equal to four times the bandwidth (that is $\approx$ 400% in terms of BW).

It is now interesting to note that, if we use the tunability range definition proposed in [5] and [22], we obtain directly:

$$\frac{f_{0,\max} - f_{0,\min}}{BW} \times 100 \approx 400\% \tag{4}$$

It is important to note that our definition of tunability does not intend to define the quality of a filter. The objective of the definition is simply to measure the tuning range in terms of the number of bandwidths. This second definition is, therefore, the one we will use in the reminder of this paper since we believe that it is more appropriate for tunable filters with constant bandwidth.

The procedure we have followed to actually design the filter is based on the group delay computation described in [6]. A lumped element model is used to obtain the reference values for the different stages. The dimensions of the real structure are obtained through optimization in a stage-by-stage procedure using CST Studio Suite (v.2022, Dassault Systémes) [38] in low precision (and without considering mechanization effects). The precision has been controlled by adjusting the computational variables, namely, the threshold of the convergence criteria, the number of passes/checks and the number of cells per max model box. Finally, with the space mapping technique described in [39], the final dimensions of the real structure are obtained using CST Studio Suite (v.2022, Dassault Systémes) for the high precision full-wave electromagnetic simulations.

Fig. 6 shows the 2D side view of the structure of the proposed filter. The final prototype consists of a single housing of dimensions $W \times H \times l$, with $W = 40\,\text{mm}$, $H = 50\,\text{mm}$ and $l = 200\,\text{mm}$, being W the width of the waveguide housing (transversal dimension not shown in Fig. 6). The resonators are located at $l_1 = 50\,\text{mm}$ and $l_2 = 50\,\text{mm}$. The height values of the hollow metal supports for the dielectric resonators and coupling structures are $h_1 = h_4 = 26.819\,\text{mm}$, $h_2 = h_3 = 30.651\,\text{mm}$, $h_i = 26.952\,\text{mm}$, $hc_1 = hc_3 = 19.681\,\text{mm}$, $hc_2 = 13.351\,\text{mm}$, respectively. All four hollow metal supports for the resonators have the same outer and inner diameter of $D_o = 16\,\text{mm}$ and $D_i = 12\,\text{mm}$, respectively. The outer diameter of all hollow metal

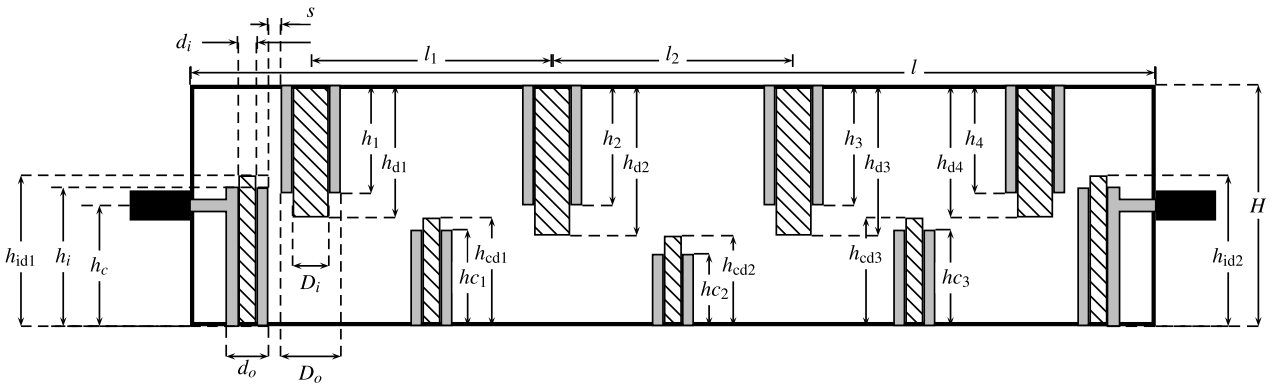


FIGURE 6. Side view of the filter structure. The filter is symmetric and the most relevant dimensions used in the design are shown.

supports for the coupling posts is $d_0 = 8$ mm, and the inner diameter is $d_i = 6$ mm. The gap s has been set to 2.0 mm to facilitate manufacturing. The coaxial connector (of external and internal diameters equal to 4.06 mm and 1.28 mm, respectively, with dielectric constant 2.1) is placed at a height of $h_c = 24.528$ mm for the input/output excitation.

It is important to point out at this stage that the starting values for the resonators, input and output couplings and inter-resonator couplings are simply obtained by optimizing the height of hollow metallic support cylinders, while keeping the dielectric rods inside the support structures. The Alumina dielectric rods, with a dielectric constant of 9.9 and loss tangent of 0.0002, are then pushed out of the hollow support posts in order to tune the filter at a lower center frequency. For this purpose, the diameters of the tuners used in the resonator and coupling structures are 11.8 mm and 5.95 mm, respectively. They are slightly smaller than the inner diameters of the corresponding hollow metal supports.

The two extreme cases for the tuning range have been obtained with the penetrations of the Alumina tuners shown in Table 1. The simulated filter performance for these two cases is shown in Fig. 7. The inset in Fig. 7 shows a detailed view of the insertion losses at the lower and upper end of the tuning range. The structure has been simulated considering the losses due to aluminum for all metal parts, and the losses for the dielectric tuning rods (Alumina). The simulations have been carried out with two different full-wave 3D simulators, namely, HFSS (Ansoft) and CST Studio Suite, in order to validate the accuracy of the results obtained. Moreover, the quality factor has been estimated using the simulation results including all losses.

As we can see, the proposed structure maintains the same filter characteristics for a tuning range of 300 MHz. The insertion loss is better than 0.06 dB when the filter is tuned at $f_{0,\min} = 1542$ MHz, giving an estimated quality factor approximately equal to 5982. When the response is tuned at $f_{0,\max} = 1842$ MHz, on the other hand, the insertion losses are better than 0.05 dB with a quality factor approximately equal to 7430.

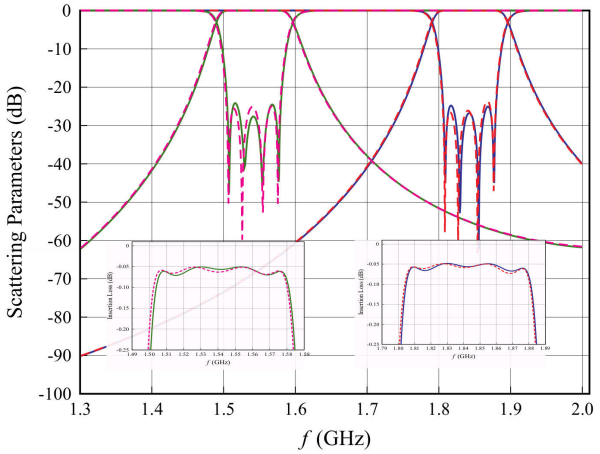


FIGURE 7. Electrical response of the filter for two extreme cases of the tuning range (the upper frequency case requires the lowest penetrations of the tuners and the lower frequency case requires the deepest penetrations). Solid lines: simulation with CST. Dashed lines: simulation with HFSS (Insertion loss response in corresponding inset figure).

III. ZIRCONIA VS ALUMINA TUNERS

In the previous section, we have shown how the structure that we propose can effectively be used to build combline filters that can be tuned in center frequency using tuning rods made of Alumina. In this section we explore the use of an alternative dielectric, namely, Zirconium Dioxide ($Y_2O_3 - ZrO_2$), or simply Zirconia. This material has, in fact, already been proposed for a number of microwave applications.

In [40], for instance, the performance of Zirconia thick films in high Q structures has been studied. Moreover, [41] has carried out comparative studies of Yttrium-stabilized Zirconia and Alumina coatings. Furthermore, the use of Zirconia in wideband filter applications has also been recently shown in [42]. The Zirconia we have use has the dielectric constant $\epsilon_r = 28$, and loss tangent $\tan\delta = 0.001$.

A. RESONATOR RESULTS

Having changed the dielectric material, we need also to redesign the basic resonator. The dimensions of the center

TABLE 1. Simulated dielectric tuners penetration (in mm).

Frequency (GHz)	h_{id1}	h_{d1}	h_{cd1}	h_{d2}	h_{cd2}	h_{d3}	h_{cd3}	h_{d4}	h_{id2}
1.842 GHz	25 mm	24 mm	16.5 mm	29 mm	12 mm	29 mm	16.5 mm	24 mm	25 mm
1.542 GHz	44.56 mm	45.07 mm	49.9 mm	47.75 mm	18 mm	47.75 mm	49.9 mm	45.07 mm	44.56 mm

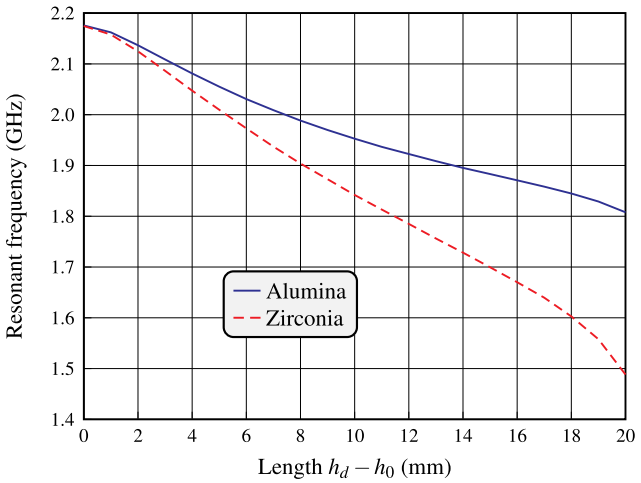


FIGURE 8. Resonant frequency in terms of the penetration of the dielectric rod exceeding the hollow post ($h_d - h_0$) according to Fig. 1.

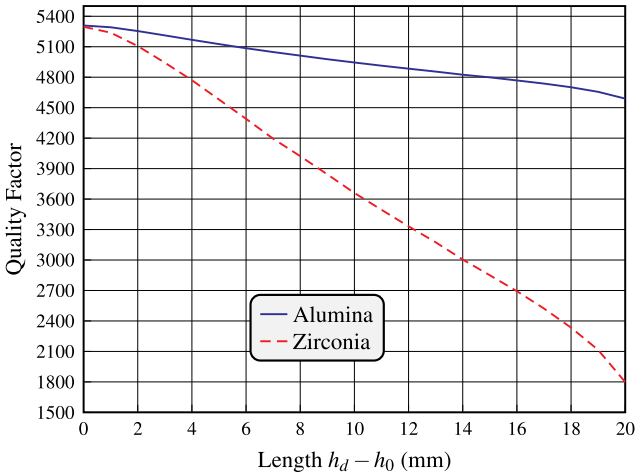


FIGURE 9. Q_u factor in terms of the penetration of the dielectric rod exceeding the hollow post ($h_d - h_0$) according to Fig. 1.

resonator post are calculated following the same procedure already discussed. The resulting cavity dimensions (shown in Fig. 1) are $l_0 = 44.8$ mm, $H = 48$ mm and $W = 35$ mm. The hollow resonator post is set to $h_0 = 27.5$ mm in order to place the initial resonant frequency (the highest one of the new design) at $f_0 = 2.175$ GHz.

One point that is now important is to make sure that the use of a higher dielectric constant does not introduce radiation loss. Radiation could, in fact, occur because the metallic posts loaded with Zirconia are actually sections of circular waveguide. If their cutoff is near the resonant frequency of the cavity, they could effectively extract energy from the filter. The diameter that we have used for the dielectric rod is $D = 9.91$ mm. The hollow post has an inner diameter $D_i = 10$ mm. The resulting cutoff frequency is 3.39 GHz, which is well above the maximum operating frequency of our filter. Another important consideration in this context is that all the hollow posts are located along the center line of the filter. As a consequence, due to symmetry, the excitation of the modes inside the hollow posts is theoretically reduced to zero. The above considerations confirm that the structure we have chosen does not suffer from losses due to radiation.

One advantage introduced by the use of Zirconia is that the tuning range is considerably extended with respect to Alumina, as shown in Fig. 8. However, a significant effect in the unloaded quality factor is observed, as shown in Fig. 9.

B. RESULTS FOR INTER RESONATOR COUPLING

In this new design, we have used a different approach to fix the initial inter-resonator coupling. The correct coupling value has, in fact, been obtained optimizing the distance between the consecutive resonator posts (l_1 and l_2 variables in Fig. 6), rather than the height of the hollow inter-resonator post. However, a hollow metallic post of $= 5$ mm in height has been maintained to provide mechanical support for the dielectric rods. An additional advantage of using Zirconia is that it provides a much stronger (and relatively linear) variation of the coupling coefficient, as compared to the one in Fig. 5. To simplify the manufacturing, the outer and inner diameters of the metallic posts are chosen to be $D_0 = 12$ mm and $D_i = 10$ mm (the same as in the resonators). A distance between two resonator posts of $l_1 = 44$ mm is, in this case, required to obtain the initial coupling coefficient value equal to 0.0185.

Furthermore, a dielectric rod of the diameter $D = 9.91$ mm is inserted inside the hollow post. It is important to note that, for the sake of comparison, we have used for this study the two types of dielectric rods (Zirconia and Alumina) with the same dimensions. The results of this experiment clearly show significantly wider coupling range values with Zirconia, as shown in Fig. 10.

C. RESULTS FOR INPUT/OUTPUT COUPLING

We re-designed next the input coupling structure in Fig. 3. The resulting dimensions are $H = 48$ mm, $l_0 = 44$ mm and $W = 35$ mm. The height of the hollow resonator post and the input coupling structure are $h_0 = 24.2$ mm and $h_i = 22.35$ mm, respectively. The coaxial input probe has been placed at the height $h_c = 21$ mm. The hollow post now

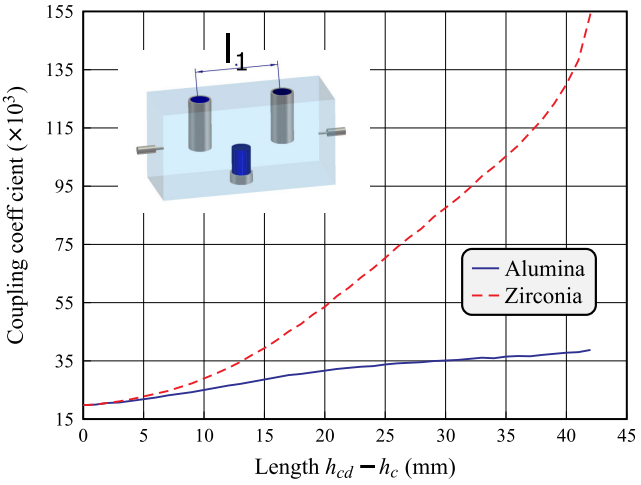


FIGURE 10. Coupling coefficient variation for two possible cases: experiment with Alumina (solid line) and the proposed solution using Zirconia (dashed line). Distance l_1 in the inset has been moved to obtain the initial coupling value at the left of the graph. Both coupling coefficients are plotted versus the penetration of the dielectric tuner exceeding the hollow post ($h_{cd} - h_c$).

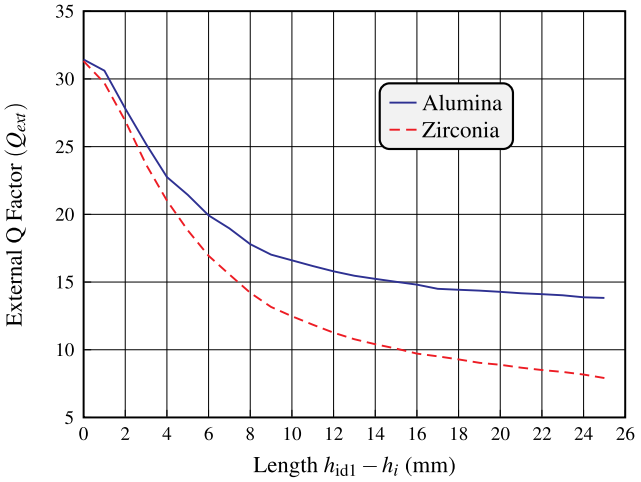


FIGURE 11. Q_{ext} variation for two possible cases: experiment with Alumina (solid line) and the proposed solution inserting Zirconia (dashed line). Both curves are plotted versus the penetration of the dielectric rod exceeding the hollow post ($h_{id1} - h_i$).

has outer and inner diameters equal to $d_o = 12$ mm and $d_i = 10$ mm, respectively. The dielectric rods of Alumina and Zirconia are inserted to explore the variation in the input coupling value. A gap s equal to 2 mm has been left between the coupling post and the resonator post.

In order to start the experiment, the dielectric rod is totally confined inside the metallic hollow post having height equal to $h_{id1} = 22$ mm. The simulations we have performed indicates that the structure provides a $Q_{ext} = 32.8$. This is indeed the value required to design the initial case (highest center frequency of 2.175 GHz and lower bandwidth of 50 MHz). As expected, Zirconia can achieve a significantly wider range of Q_{ext} with respect to the h_{id1}, as shown in Fig. 11.

IV. TUNABLE FILTER IN FREQUENCY AND BANDWIDTH

Using the information obtained with respect to the behavior of Zirconia, we have next re-designed the filter structure using the same procedure already described for the Alumina case in Fig. 6. The novel prototype consists of a single housing (see Fig. 6) of dimensions $W \times H \times l$ with $W = 35$ mm, $H = 48$ mm and $l = 177.8$ mm.

As previously discussed in section III-B, in this case we have optimized the length between consecutive resonators to obtain the initial values for the inter resonator couplings. This leads to the initial values $l_1 = 41.82$ mm and $l_2 = 46.12$ mm. The height of the hollow resonator posts are set to $h_1 = h_4 = 24.45$ mm and $h_2 = h_3 = 26.82$ mm. All inter-resonator coupling hollow posts have equal height of $hc_1 = hc_2 = hc_3 = 5$ mm that also act as the support structures for the dielectric tuners. In order to simplify the structure, all the hollow metallic posts used for either inter resonator couplings and input/output couplings are set to the same outer diameter of $d_o = D_o = 12$ mm. All inner diameters are set to $d_i = D_i = 10$ mm, which allows the same size dielectric rod ($D = 9.91$ mm) to be utilized for all the tuning elements.

The input coupling post having the height of $h_i = 22.37$ mm has been excited with a coaxial probe at the height of $h_c = 21$ mm. The gap s has been set to 2mm to simplify manufacturing. The coaxial connector for the input and output excitation is the same used in section II.

In order to provide an accurate prediction of the filter performance, the structure has been simulated taking into account the metal losses in the aluminum for the hollow posts and the housing. Dielectric losses are also taken into account in all rods used for the resonators and for the coupling structures. It is important to note that the filter structures discussed in this paper are reconfigurable also in bandwidth in addition to center frequency. For this last filter, for instance, the lower and upper limits of the bandwidth are $BW = 50$ MHz and $BW = 100$ MHz, respectively. This means that the filter can be tuned to achieve any bandwidth value between 50 MHz to 100 MHz throughout the complete center frequency tuning range while maintain an equiripple return loss $RL > 25$ dB in all cases. The specifications for the initial design of this fully reconfigurable filter are as follows:

Case 1:

- Center frequency must be tunable from $f_{0,\min} = 1.475$ GHz up to $f_{0,\max} = 2.175$ GHz.
- Bandwidth $BW = 50$ MHz constant throughout the whole range of frequencies.
- Return losses: $RL > 25$ dB.

and must be achieved while maintaining a constant 50MHz bandwidth and return loss performance. This means that the filter tuning range (as defined in [5] and [22]) is

$$\frac{f_{0,\max} - f_{0,\min}}{BW} \times 100 = 1400\% \tag{5}$$

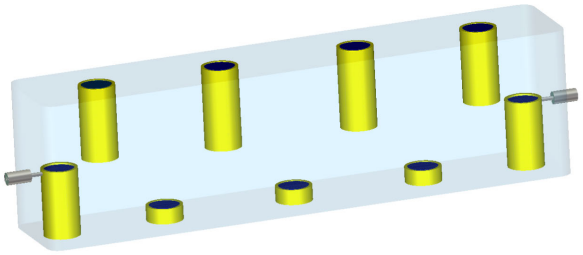


FIGURE 12. 3-D model of the combline filter with tunable frequency and bandwidth.

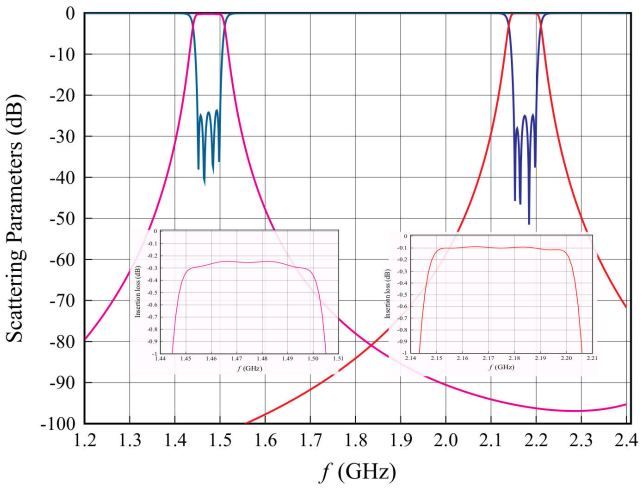


FIGURE 13. Simulation response of the filter for two extreme cases of the tuning range (the upper frequency case requires the lowest penetrations of the tuners, and the lower frequency case requires the deeper penetrations) while keeping the initial 50 MHz bandwidth. Blue line: Simulated S_{11} at 2175 MHz. Red line: Simulated S_{21} at 2175 MHz. Midnight Blue line: Simulated S_{11} at 1475 MHz. Magenta line: Simulated S_{21} at 1475 MHz (Insertion loss in corresponding inset figure).

Case 2:

- Center frequency must be tunable from $f_{0,\min} = 1.60\,\text{GHz}$ up to $f_{0,\max} = 1.92\,\text{GHz}$.
- Bandwidth $BW = 100\,\text{MHz}$ constant throughout the whole range of frequencies.
- Return losses: $RL > 25\,\text{dB}$.

This means that the filter tuning range (as defined in [5] and [22]) is

$$\frac{f_{0,\max} - f_{0,\min}}{BW} \times 100 = 320\% \tag{6}$$

Fig. 12 shows the final structure of the filter. As we can see, for case 2 ($BW = 100\,\text{MHz}$) the frequency tuning range is smaller than for case 1 ($BW = 50\,\text{MHz}$). This is due to the loading effect of the dielectric tuners, which need to be pushed out of the hollow posts in the coupling elements to obtain a wider bandwidth ($BW = 100\,\text{MHz}$), thus reducing the maximum frequency value that can be achieved. The penetration of the Zirconia tuners is summarized in Table 2 for the lower and higher cases of the lowest (50 MHz) and highest (100 MHz) filter bandwidths.

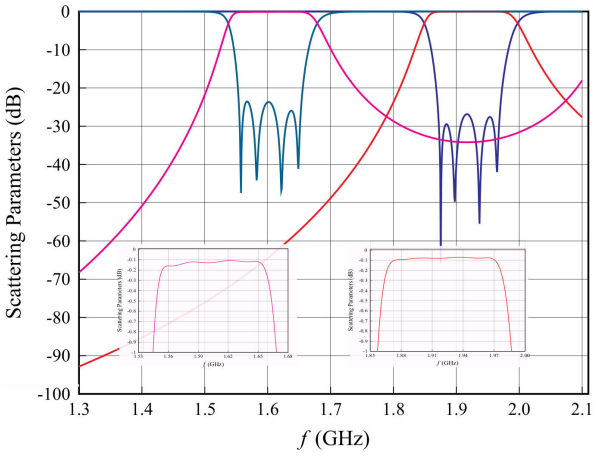


FIGURE 14. Simulation response of the filter for two extreme cases of the tuning range (the upper frequency case requires the lowest penetrations of the tuners, and the lower frequency case requires the deeper penetrations) while keeping the 100 MHz bandwidth. Blue line: Simulated S_{11} at 1920 MHz. Red line: Simulated S_{21} at 1920 MHz. Midnight Blue line: Simulated S_{11} at 1600 MHz. Magenta line: Simulated S_{21} at 1600 MHz (Insertion loss response in corresponding inset figure).

Fig. 13 shows the simulated performance obtained by changing the filter center frequency while maintaining a constant bandwidth of 50 MHz. Fig. 14, on the other hand, shows the simulated performance obtained maintaining a constant bandwidth of 100 MHz. It is worth noting at this stage that our simulations predicted the in-band insertion loss of 0.3 dB for 50 MHz and 0.25 dB for 100 MHz. These theoretical values are in good agreement with the experimental results of the manufactured prototype shown next.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The manufacturing drawings of the filter prototype have been obtained using SolidWorks (from Dassault Systémes SolidWorks Corp.). The structure is composed of one body and two covers. Fig. 15 shows the 3D model of the filter.

As shown in Fig. 15, an additional 10 mm length of hollow metallic post is added to the external surface of both top and bottom covers. This has been done to provide mechanical support to the dielectric rods. Furthermore, a Teflon screw has been added in the orthogonal direction, to hold the dielectric rods in the desired tuning position. It is important to note at this point that, due to the use of milling, all concave corners of the filter body now have a curvature radius equal to 2 mm. The filter is made in Aluminum without silver plating.

Fig. 16 shows the internal topology of the filter before assembly. Furthermore, Fig. 17 shows the prototype of the filter assembles using dielectric tuners made of Zirconia. The filter has been tested using a vector network analyzer (VNA), after being properly calibrated following a standard TRL procedure. The filter is initially tuned at the higher center frequency of 2175 MHz with the lowest $BW = 50\,\text{MHz}$. Then, as it has already been discussed, the bandwidth of the

TABLE 2. Simulated Zirconia tuners penetration (in mm).

Frequency (GHz)	Bandwidth (MHz)	h_{id1}(mm)	h_{d1}(mm)	h_{cd1}(mm)	h_{d2}(mm)	h_{cd2}(mm)	h_{d3}(mm)	h_{cd3}(mm)	h_{d4}(mm)	h_{id2}(mm)
2.175 GHz	50 MHz	22	24	5	26.4	5	26.4	5	24	22
1.475 GHz	50 MHz	30.10	44.60	15.77	46.8	13.93	46.8	15.77	44.60	30.10
1.920 GHz	100 MHz	32.28	23.86	30.63	30.94	27.65	30.94	30.63	23.86	32.28
1.600 GHz	100 MHz	41.68	31.28	46.99	41.06	44.62	41.06	46.99	31.28	41.68

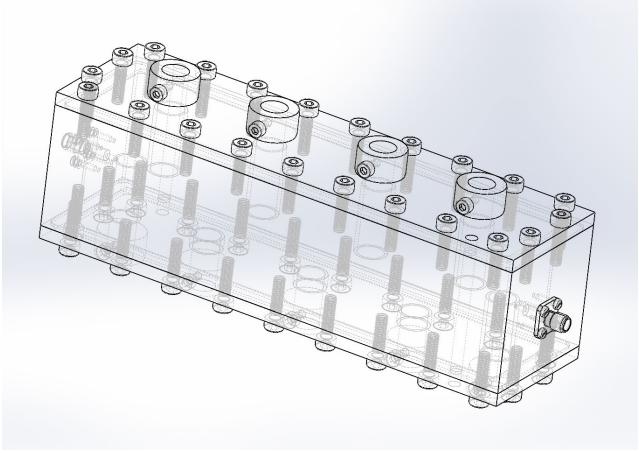


FIGURE 15. Final structural model (developed with SolidWorks) of the filter prototype.

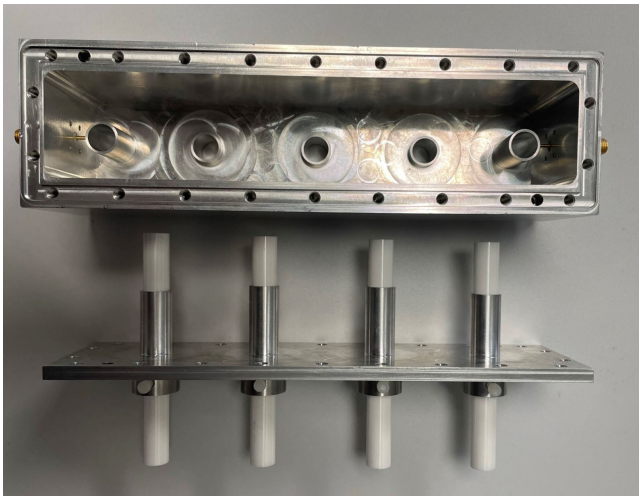


FIGURE 16. Filter structure before assembly.

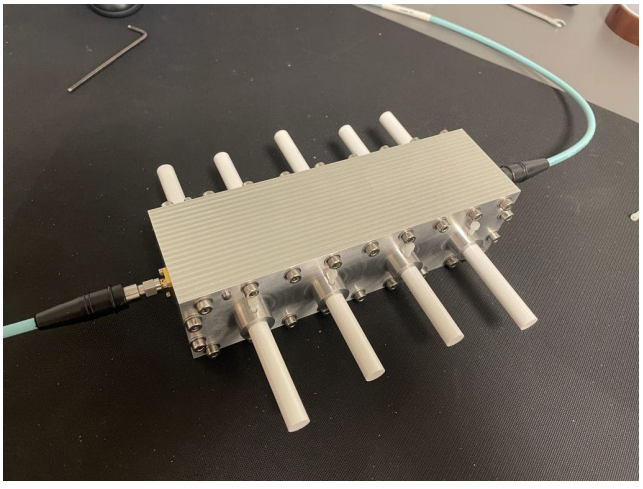


FIGURE 17. Final assembled filter prototype on the test bench.

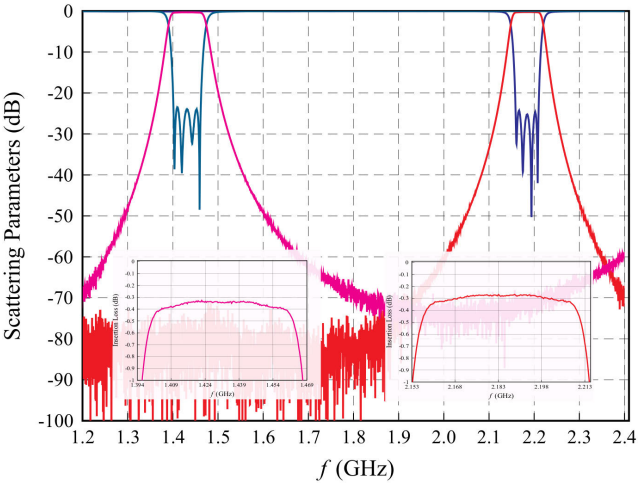


FIGURE 18. Measured response of the filter for two extreme cases of the tuning range while keeping the initial 50 MHz bandwidth. Blue line: Measured S_{11} at 2180 MHz. Red line: Measured S_{21} at 2180 MHz. Midnight Blue line: Measured S_{11} at 1433 MHz. Magenta line: Measured S_{21} at 1433 MHz (Insertion loss response in corresponding inset figure).

bandpass response can be tuned to any value between 50 MHz and 100 MHz, in the corresponding frequency ranges. As a proof of concept, the two extreme cases (i.e. $BW = 50$ MHz and $BW = 100$ MHz) have been demonstrated. In the first case, the filter center frequency can be tuned while keeping a constant bandwidth equal to 50 MHz, as shown in Fig. 18. The achieved frequency tuning range is 1494%, namely, from 1433 MHz to 2180 MHz. The response of the filter has an insertion loss level better than 0.35 dB throughout the complete tuning range, as shown in the inset of Fig. 18.

The next measurement has been performed expanding the bandwidth up to 100 MHz (double of the initial bandwidth). In this case, the filter can achieve a tuning range of 320%, starting from 1600 MHz down to 1920 MHz. As shown in Fig. 19, the measured results show a return loss level always better than 25 dB, and insertion loss values below 0.22 dB through the whole tuning range.

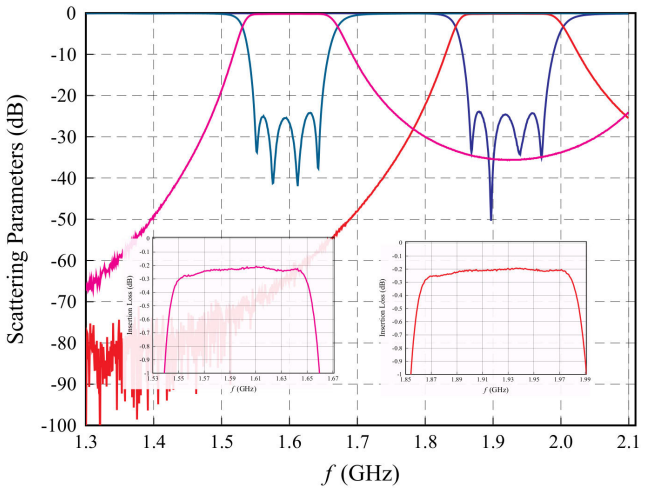


FIGURE 19. Measured response of the filter for two extreme cases of the tuning range while keeping the initial 100 MHz bandwidth. Blue line: Measured S_{11} at 1920 MHz. Red line: Measured S_{21} at 1920 MHz. Midnight Blue line: Measured S_{11} at 1600 MHz. Magenta line: Measured S_{21} at 1600 MHz (Insertion loss response in corresponding inset figure).

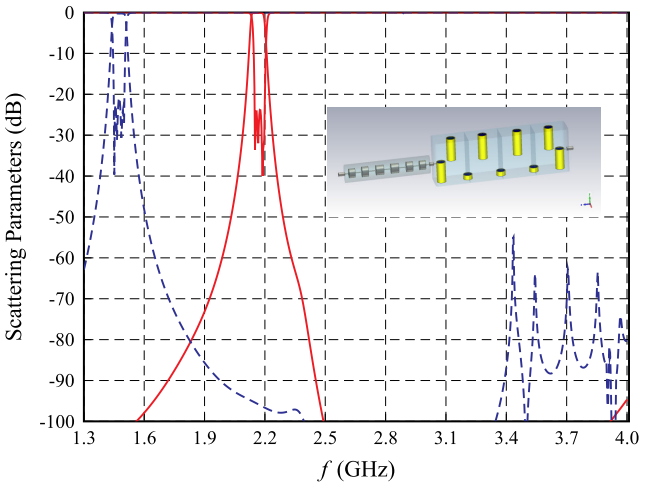


FIGURE 20. Simulated response of the integrated structure with LPF for two extreme cases of the tuning range while keeping the initial 50 MHz bandwidth. Red line: Simulated response at 2175 MHz. Blue dashed line: Simulated response at 1475 MHz. (Lowpass filter + Bandpass filter integrated structure in corresponding inset figure).

One important aspect to evaluate the performance of our filter structure is the achieved quality factor. To this end, the Q factor has been evaluated and compared using the insertion loss of the measured and simulated responses [43]. The results obtained are shown in Table. 3. The frequency range used is $1475 - 2175$ MHz. The extracted Q factor from measurement appears to be in agreement with the simulated estimations. It is important to note at this point that the prototype has been manufactured using aluminum. The Q factor of the filter can be further improved using silver plating, as can be seen in [44].

VI. OUT-OF-BAND PERFORMANCE

One of the key characteristics of any microwave filter is the out-of-band performance. However, the specific requirements that a filter must satisfy are normally dictated by the intended application. In this context, therefore, since the main objective of this paper is to discuss the proof-of-concept for a filter structure that can be reconfigured both in terms of center frequency and bandwidth, we are not aiming at satisfying a specific set of requirements. Nevertheless, since the out-of-band performance is indeed an important aspect, we discuss in this section a number of examples of how the spurious performance of our filter can be easily controlled with the help of simple, standard lowpass filters (LPFs).

Generally speaking, the out of band performance of a filter is limited by the appearance of spurious responses that can be due to the harmonics of the resonators, or to the excitation of higher order modes. For the reconfigurable filter discussed in this paper, the out-of-band response is intrinsically linked to the filter bandwidth. In fact, increasing the bandwidth, will inevitably reduce the spurious free range above the filter passband.

As a first illustrative example, we have designed a simple six-pole stepped-impedance coaxial lowpass filter, with a cutoff frequency of 2.3 GHz [45]. Fig. 20 shows the structure of the LPF integrated with the reconfigurable filter, and the simulated response for the case in which the bandwidth is set to the minimum value of 50 MHz. As we can see, adding a simple LPF of sixth order, we have obtained more than 60 dB of rejection up to 4 GHz.

As a second example, we have set the bandpass of our reconfigurable filter to 100 MHz, and we have designed an eight-pole coaxial LPF with a cutoff frequency of 2.02 GHz. Fig. 21 shows the structure and the simulated response. As expected, the LPF does improve the out-of-band response.

As a final example, we have demonstrated the combined case of an integrated lowpass filter with all the frequency and bandwidth scenarios. In this example, we have varied the bandwidths from 50 to 100 MHz. In this case, we have used an eight-pole LPF with a cutoff frequency set at 2.23 GHz. As we can see from Fig. 22, also in this case, a simple LPF can be used to improve the out-of-band response. In this case, however, the minimum frequency to which the filter can be tuned is limited to 1750 MHz, instead of 1600 MHz for 100 MHz bandwidth, as shown in Fig. 22.

The dimensions of the LPFs used in these three illustrations are given in Table 4. In all the lowpass filters discussed, the internal air spaces are filled with Teflon (relative permittivity value of 2.2). The input and output connector dimensions are identical to the ones of the filter described in subsection II-D.

VII. MULTIPACTION ANALYSIS

The power handling capability of microwave filters is a key issue for many applications and, as it is well-known, it must be evaluated in detail for each specific new structure that

TABLE 3. Q-values extracted from measurements and simulations.

Bandwidth (MHz)	Q-values extracted from measurements			Q-values extracted from simulations		
——	Frequency	Insertion loss at f_0(dB)	Q factor	Frequency	Insertion loss at f_0(dB)	Q factor
50 MHz	2180 MHz	0.320	2476	2175 MHz	0.102	7450
50 MHz	1433 MHz	0.341	1526	1475 MHz	0.255	2021
100 MHz	1920 MHz	0.202	1776	1920 MHz	0.079	4245
100 MHz	1600 MHz	0.215	1387	1600 MHz	0.129	2167

TABLE 4. Lowpass filter dimentions (in mm).

Cutoff Freq. (GHz)	l_{01}	l_{02}	l_{03}	l_{04}	l_{05}	l_{06}	l_{07}	l_{08}	l_{09}	R_{01}	R_{02}	R_{03}	R_{04}	R_{05}	R_{06}	R_{07}	R_{08}	R_{09}
2.3	0.647	7.792	9.186	8.224	10.34	9.973	9.755	——	——	1.968	5.062	0.593	5.568	0.619	5.358	0.552	——	——
2.02	0.519	5.989	10.81	8.227	10.18	9.614	8.718	10.82	8.863	1.392	5.832	0.646	6.143	0.443	6.016	0.290	5.876	0.311
2.23	0.535	5.022	9.169	7.203	9.359	8.067	7.324	9.427	7.759	0.942	5.784	0.519	5.991	0.422	5.936	0.222	5.767	0.273

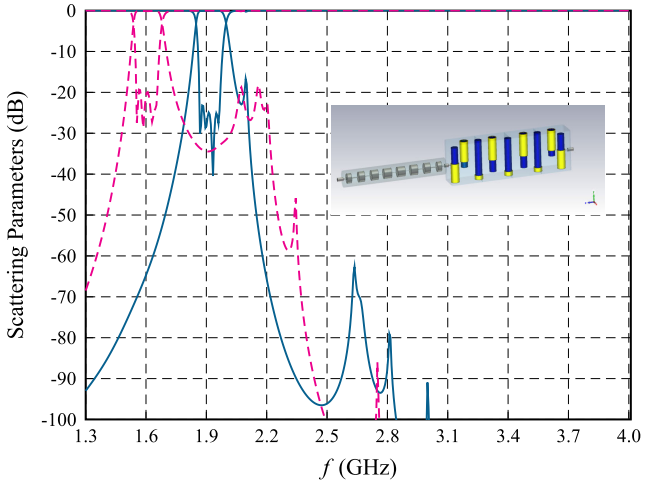


FIGURE 21. Simulated response of the integrated structure with LPF for two extreme cases of the tuning range while keeping the initial 100 MHz bandwidth. Midnight Blue line: Simulated response at 1920 MHz. Magenta dashed line: Simulated response at 1600 MHz. (Lowpass filter + Bandpass filter integrated structure in corresponding inset figure).

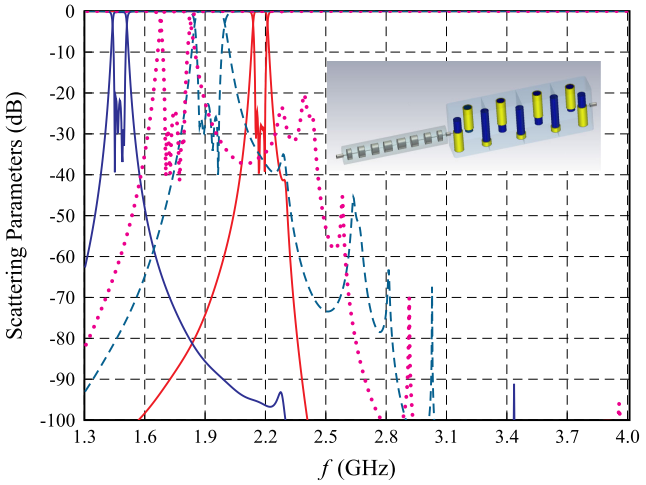


FIGURE 22. Simulated response of the integrated structure with LPF for extreme cases of the tuning range. Red line: Simulated response at 2175 MHz. Blue line: Simulated response at 1475 MHz. Midnight Blue dashed line: Simulated response at 1920 MHz. Magenta dotted line: Simulated response at 1750 MHz. (Lowpass filter + Bandpass filter integrated structure in corresponding inset figure).

is being developed. We, therefore, discuss in this section the results of a multipactor breakdown analysis for our reconfigurable filter [46]. It is important to note, however, that multipaction discharge can only happen in vacuum. As a consequence, we do not anticipate any power restrictions for ground applications.

To continue, it is important to note that, early in the design process of any filter, it is critical to evaluate and assess the power-handling phenomena of the structure being designed. In this context, the following are some key considerations relevant to multipaction:

- Coupling windows are very often critical for multipactor breakdown. As a consequence, structures without coupling windows can improve the power performance.
- The critical areas of the filter structure, namely, where the maximum electric field is located, must be identified so that a minimum gap can be used to avoid breakdown [47].
- Furthermore, since our filter structure uses both aluminum and dielectrics, the secondary emission yield (SEY) profile of the dielectric material is assumed to be the same as that of aluminum [48].
- The multipaction analysis requires the precise evaluation of the electromagnetic fields in the critical areas. The critical areas must therefore be identified, and a dense local meshing must be used to ensure the accuracy of the simulations using CST Studio Suite.

To continue, the parameters we used for the multipactor analysis, as required by the ECSS-E-HB-20-01A– Multipactor handbook [49], are as follows:

A. ALUMINUM

- SEY_{max}: 2.92.
- SEY_0: 0.8.
- E_1(eV), Impact energy: 17 eV.
- E_{max}(eV), Impact energy @ SEY_{max}: 276 eV.

TABLE 5. Simulated multipaction threshold comparison (in Watt).

Frequency Case (MHz)	Bandwidth (MHz)	Peak Power (W) @f_{min}		Peak Power (W) @f_0		Peak Power (W) @f_{max}	
2175	50	3843.62	@2.15 GHz	3656.12	@2.175 GHz	4031.12	@2.2 GHz
1720	50	137.69	@1.695 GHz	240.23	@1.72 GHz	496.08	@1.745 GHz
1670	50	90.82	@1.645 GHz	147.46	@1.67 GHz	110.35	@1.695 GHz
1620	50	46.38	@1.595 GHz	76.66	@1.62 GHz	55.17	@1.645 GHz
1571	50	19.67	@1.546 GHz	31.00	@1.571 GHz	24.17	@1.596 GHz
1545	50	16.23	@1.52 GHz	19.41	@1.545 GHz	17.21	@1.57 GHz
1475	50	4.55	@1.45 GHz	5.92	@1.475 GHz	7.87	@1.5 GHz
Frequency Case (MHz)	Bandwidth (MHz)	Peak Power (W) @f_{min}		Peak Power (W) @f_0		Peak Power (W) @f_{max}	
1920	100	1984.31	@1.87 GHz	1734.3	@1.92 GHz	1984.31	@1.97 GHz
1645	100	145.5	@1.595 GHz	205.07	@1.645 GHz	244.13	@1.695 GHz
1600	100	40.77	@1.55 GHz	120.11	@1.6 GHz	165.04	@1.65 GHz

B. POWER LOOP

- Precision: 0.1 dB.
- Initial Power: 500 W.
- Maximum Power: $1e + 06$ W.
- The initial number of electrons: 2000.

The power threshold has been estimated using SPARK3D v2022 (Aurora Software and Testing S.L.U., CST/3DS). The frequencies at which the multipaction threshold have been evaluated are three, namely, the center frequency and the ones at the edges of the bandwidth. The maximum peak power that we obtained for different cases in the tuning range is shown in Table 5. As we can see in Table 5, the filter structure is able to support a continuous wave of around 3656.12 W at the highest center frequency, as shown in row 1 of Table 5 for a 50 MHz bandwidth. However, as expected, there is significant degradation of power handling in the lower tuning frequency of 1475 MHz. When the filter is tuned to a bandwidth of 100 MHz, a similar pattern can be observed, as shown in the second section of Table 5.

It is therefore evident that there is a relation between the maximum power and the minimum tuning frequency. As an illustrative example, for instance, if we assume that the maximum required power is 100 W CW, the filter can be tuned from 2175 MHz to 1720 MHz (for 50 MHz bandwidth) and from 1920 MHz to 1645 MHz (for 100 MHz bandwidth).

Finally, it is important to note that, if higher power and a wider tuning range are required, enclosure of the filter structure in a pressurized environment would allow to use the full tuning range that the structure can offer.

VIII. REMOTE FILTER OPERATION

One important feature of the family of filters described in this paper, is that the use of dielectric tuners opens the possibility of developing remotely operated reconfigurable filters based on the use of low-cost, computer-controlled linear actuators.

In particular, this feature can be appreciated noting that, using the data in Table 2 for the central dielectric coupling rod movements, we obtain 13 μm/MHz for the 50 MHz bandwidth, and 57 μm/MHz for the 100 MHz bandwidth case.

These values are indeed perfectly compatible with commercial computer-controlled, low-cost linear actuators. Furthermore, it is important to note that, even though each tuner needs to be operated independently, developing the appropriate software, the complete filter performance can be controlled with a single command.

IX. CONCLUSION

In this paper, we have explored novel configurations to develop fully reconfigurable combline filters. The solution that we propose is based on the use of dielectric tuners. All resonators and all couplings can be tuned independently. The main advantage of using dielectrics is that it greatly reduces the sensitivity of the filter performance with respect to the position of the tuners. As a result, the use of a low-cost computer-controlled linear actuator becomes now an attractive possibility for the implementation of remotely reconfigurable combline filters. The out-of-band performance of our filter is also discussed, showing how a simple low-pass filter can be used to satisfy specific requirements. Furthermore, the multipaction power threshold of the structure has been studied, clearly showing that there is a trade-off between the maximum power and the minimum tuning frequency in vacuum. In conclusion, with the results presented in this paper, we have firmly established the proof of concept of a novel family of combline filters that is reconfigurable over a wide range of both bandwidth and center frequency. It is our opinion that the structure that we propose, may prove to be a significant contribution to the development of future reconfigurable payloads for both ground and space applications.

REFERENCES

[1] A. L. Allison, *The ITU and Managing Satellite Orbital and Spectrum Resources in the 21st Century* (SpringerBriefs in Space Development). Cham, Switzerland: Springer, 2014.

[2] G. Varrall, *Making TelecomsWork: From Technical Innovation to Commercial Success*. Chichester, U.K.: Wiley, 2012.

[3] J.-S. Hong, "Reconfigurable planar filters," *IEEE Microw. Mag.*, vol. 10, no. 6, pp. 73–83, Oct. 2009.

[4] R. R. Mansour, F. Huang, S. Fouladi, W. D. Yan, and M. Nasr, "High-*Q* tunable filters: Challenges and potential," *IEEE Microw. Mag.*, vol. 15, no. 5, pp. 70–82, Jul. 2014.

[5] E. Laplanche, N. Delhote, A. Périgaud, O. Tantot, S. Verdeyme, S. Bila, D. Pacaud, and L. Carpentier, "Tunable filtering devices in satellite payloads: A review of recent advanced fabrication technologies and designs of tunable cavity filters and multiplexers using mechanical actuation," *IEEE Microw. Mag.*, vol. 21, no. 3, pp. 69–83, Mar. 2020.

[6] R. J. Cameron, C. M. Kudsia, and R. R. Mansour, *Microwave Filters for Communication Systems: Fundamentals, Design, and Applications*, 2nd ed. Hoboken, NJ, USA: Wiley, 2018.

[7] M. Yu, B. Yassini, B. Keats, and Y. Wang, "The sound the air makes: High-performance tunable filters based on air-cavity resonators," *IEEE Microw. Mag.*, vol. 15, no. 5, pp. 83–93, Jul. 2014.

[8] C. Arnold, J. Parlebas, R. Meiser, and T. Zwick, "Fully reconfigurable manifold multiplexer," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3885–3891, Oct. 2017.

[9] B. Yassini, M. Yu, D. Smith, and S. Kellett, "A *Ku*-band high-*Q* tunable filter with stable tuning response," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 12, pp. 2948–2957, Dec. 2009.

[10] S. Nam, B. Lee, C. Kwak, and J. Lee, "Contactless tuning plunger and its application to K-band frequency-tunable cavity filter," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 2713–2719, Jul. 2019.

[11] L. Pelliccia, F. Cacciamani, P. Farinelli, and R. Sorrentino, "High-*Q* tunable waveguide filters using ohmic RF MEMS switches," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 10, pp. 3381–3390, Oct. 2015.

[12] U. Rosenberg, R. Beyer, P. Krauß, T. Sieverding, A. Papanastasiou, M. Pueyo-Tolosa, P. M. Iglesias, and C. Ernst, "Novel remote controlled dual mode filter providing flexible re-allocation of center frequency and bandwidth," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 2016, pp. 1–3.

[13] A. Périgaud, O. Tantot, N. Delhote, S. Verdeyme, S. Bila, D. Pacaud, L. Carpentier, J. Puech, L. Lapierre, and G. Carayon, "Continuously tuned Ku-band cavity filter based on dielectric perturbers made by ceramic additive manufacturing for space applications," *Proc. IEEE*, vol. 105, no. 4, pp. 677–687, Apr. 2017.

[14] S. Nam, B. Lee, C. Kwak, and J. Lee, "A new class of K-band high-Q frequency-tunable circular cavity filter," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1228–1237, Mar. 2018.

[15] F. Huang and R. R. Mansour, "Tunable compact dielectric resonator filters," in *Proc. Eur. Microw. Conf. (EuMC)*, Oct. 2009, pp. 559–562.

[16] W. Dong Yan and R. R. Mansour, "Tunable dielectric resonator bandpass filter with embedded MEMS tuning elements," *IEEE Trans. Microw. Theory Techn.*, vol. 55, no. 1, pp. 154–160, Jan. 2007.

[17] S. Fouladi, F. Huang, W. D. Yan, and R. R. Mansour, "High-Q narrowband tunable combline bandpass filters using MEMS capacitor banks and piezomotors," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 393–402, Jan. 2013.

[18] A. F. Azarnaminy and R. Mansour, "A combline tunable filter with loss compensation circuit," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2018, pp. 1367–1369.

[19] F. Huang, S. Fouladi, and R. R. Mansour, "High-*Q* tunable dielectric resonator filters using MEMS technology," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 12, pp. 3401–3409, Dec. 2011.

[20] J.-X. Xu, L. Yang, Y. Yang, and X. Y. Zhang, "High-*Q*-factor tunable bandpass filter with constant absolute bandwidth and wide tuning range based on coaxial resonators," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4186–4195, Oct. 2019.

[21] J. C. Melgarejo, J. Ossorio, S. Cogollos, M. Guglielmi, V. E. Boria, and J. W. Bandler, "On space mapping techniques for microwave filter tuning," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 4860–4870, Dec. 2019.

[22] J. Ossorio, J. Vague, V. E. Boria, and M. Guglielmi, "Exploring the tuning range of channel filters for satellite applications using electromagnetic-based computer aided design tools," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 2, pp. 717–725, Feb. 2018.

[23] C. Kwak, M. Uhm, and I. Yom, "Feasibility study on combline filter for tunable filters," in *Proc. Asia–Pacific Microw. Conf. (APMC)*, Nov. 2013, pp. 927–929.

[24] J. S. Parish, N. Somjit, and I. C. Hunter, "Continuous frequency and bandwidth tunable combline cavity bandpass filters with internally mounted motors," *IET Microw., Antennas Propag.*, vol. 12, no. 15, pp. 2332–2337, Dec. 2018.

[25] E. Doumanis, G. Goussetis, J. Vuorio, K. Hautio, O. Amper, E. Kuusmik, and J. Pallonen, "Tunable filters for agile 5G new radio base transceiver stations," *IEEE Microw. Mag.*, vol. 22, no. 11, pp. 26–37, Nov. 2021.

[26] G. Basavarajappa and R. R. Mansour, "Design methodology of a tunable waveguide filter with a constant absolute bandwidth using a single tuning element," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5632–5639, Dec. 2018.

[27] R. R. Mansour, "A tunable quarter-wavelength coaxial filter with constant absolute bandwidth using a single tuning element," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 6, pp. 658–661, Jun. 2021.

[28] J. Ossorio, V. E. Boria, and M. Guglielmi, "Dielectric tuning screws for microwave filters applications," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2018, pp. 1253–1256.

[29] G. Macchiarella, L. Accatino, and A. Malagoli, "Design of Ka-band tunable filters in rectangular waveguide with constant bandwidth," in *Proc. IEEE Asia–Pacific Microw. Conf. (APMC)*, Dec. 2020, pp. 622–624.

[30] A. Sharma, J. Ossorio, D. Smacchia, T. Tronser, S. Cogollos, V. E. Boria, and M. Guglielmi, "Remote reconfiguration of microwave filters using dielectric tuners," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 7, pp. 3051–3064, Jul. 2023.

[31] M. Hoeft, A. Kronberger, and O. Bartz, "Tunable bandpass filters for multi-standard applications," in *Proc. German Microw. Conf.*, Mar. 2008, pp. 1–4.

[32] R. Beyer, S. Heisen, and U. Rosenberg, "CAD of conductor loaded cavity filters with wide band tuning capability," in *Proc. 32nd Eur. Microw. Conf.*, Sep. 2002, pp. 1–4.

[33] A. Sharma, S. Cogollos, V. E. Boria, and M. Guglielmi, "Analysis and design of re-configurable combline filters using dielectric tuners," in *Proc. 51st Eur. Microw. Conf. (EuMC)*, Apr. 2022, pp. 1–12.

[34] M. El Sabbagh, K. A. Zaki, H.-W. Yao, and M. Yu, "Full-wave analysis of coupling between combline resonators and its application to combline filters with canonical configurations," *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 12, pp. 2384–2393, Dec. 2001.

[35] Y. Xie, F.-C. Chen, and Q.-X. Chu, "Tunable cavity filter and diplexer using in-line dual-post resonators," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 6, pp. 3188–3199, Jun. 2022.

[36] A. Widaa, C. Bartlett, and M. Höft, "Tunable coaxial bandpass filters based on inset resonators," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 1, pp. 285–295, Jan. 2023.

[37] W. Lin, K. Zhou, and K. Wu, "Band-reconfigurable tunable bandpass filters based on mode-switching concept," *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 3, pp. 1125–1135, Mar. 2023.

[38] (Dec. 22, 2022). *CST Microwave Studio*. [Online]. Available: https://www.3ds.com/products-services/simulia/products/cst-studiosuite/

[39] J. Ossorio, J. C. Melgarejo, V. E. Boria, M. Guglielmi, and J. W. Bandler, "On the alignment of low-fidelity and high-fidelity simulation spaces for the design of microwave waveguide filters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5183–5196, Dec. 2018.

[40] P. A. Smith, "Performance of YBCO thick films in high-Q structures," in *Proc. IEE Colloq. Superconducting Microw. Circuits*, 1996, pp. 1–5.

[41] Y. Burachevsky and I. Bakeev, "Properties of electrical insulation coatings from yttrium-stabilized zirconia and alumina," in *Proc. 7th Int. Congr. Energy Fluxes Radiat. Effects (EFRE)*, Sep. 2020, pp. 827–830.

[42] Y. Marchives, N. Delhote, S. Verdeyme, and P. M. Iglesias, "Wide-band dielectric filter at C-band manufactured by stereolithography," in *Proc. 44th Eur. Microw. Conf.*, Oct. 2014, pp. 187–190.

[43] G. L. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*. Norwood, MA, USA: Artech House, 1980.

[44] S. Shukla, N. Gomathi, and R. George, "Autocatalytic silver-plating of aluminium radio frequency waveguides with autocatalytic nickel as the undercoat for space applications," *Surf. Topography, Metrology Properties*, vol. 2, no. 4, pp. 1–7, Nov. 2014.

[45] R. Levy and T. E. Rozzi, "Precise design of coaxial low-pass filters," *IEEE Trans. Microw. Theory Techn.*, vol. MTT-16, no. 3, pp. 142–147, Mar. 1968.

[46] E. Sorolla, S. Anza, B. Gimeno, A. M. P. Perez, C. Vicente, J. Gil, F. J. S. Prez, F. D. Quesada, A. Alvarez, and V. E. Boria, "An analytical model to evaluate the radiated power spectrum of a multipactor discharge in a parallel-plate region," *IEEE Trans. Electron Devices*, vol. 55, no. 8, pp. 2252–2258, Aug. 2008.

[47] J. de Lara, F. Perez, M. Alfonseca, L. Galan, I. Montero, E. Roman, and D. R. Garcia-Baquero, "Multipactor prediction for on-board spacecraft RF equipment with the MEST software tool," *IEEE Trans. Plasma Sci.*, vol. 34, no. 2, pp. 476–484, Apr. 2006.
[48] J. Vague, J. C. Melgarejo, M. Guglielmi, V. E. Boria, S. Anza, C. Vicente, M. R. Moreno, M. Taroncher, B. G. Martínez, and D. Raboso, "Multipactor effect characterization of dielectric materials for space applications," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 8, pp. 3644–3655, Aug. 2018.
[49] *Multipacting Handbook*, document ECSS-E-HB-20-01A, European Cooperation for Space Standarization (ECSS), ESA-ESTEC, ESA Publication Division, Amsterdam, The Netherlands, Jun. 2020.

ABHISHEK SHARMA (Student Member, IEEE) was born in New Delhi, India, in 1992. He received the bachelor's degree in electronics and communication from the Technical University of Uttar Pradesh, India, in 2014, and the master's degree in microwave and optical communication from the Department of Electronics and Communication, Delhi Technological University (formerly DCE), India, in 2017. He is currently pursuing the Ph.D. degree with the Microwave Applications Group (GAM-ITEAM), Universitat Politècnica de València (UPV), Valencia, Spain, like an Early Stage Researcher of the TESLA Network.

From 2017 to 2019, he was an RF Design Engineer with the Microwave Component Division of Bharat Electronics Ltd., under the Ministry of Defence, India, where he participated in the design of TR modules for active phased arrays and radars. During the Ph.D. studies, he developed novel compact solutions for reconfigurable waveguide filters and multiplexers. His research interests include high-power filter design, reconfigurable filters, transmit-receive modules, and RF subsystems and radars.

SANTIAGO COGOLLOS (Member, IEEE) was born in Valencia, Spain, in 1972. He received the "Ingeniero Superior de Telecomunicación" and "Doctor Ingeniero de Telecomunicación" degrees from Universitat Politècnica de València (UPV), Valencia, in 1996 and 2002, respectively.

In 2000, he joined the Communications Department, UPV, where he was an Assistant Lecturer, from 2000 to 2001; a Lecturer, from 2001 to 2002; an Associate Professor, from 2002 to 2022; and became a Full Professor, in 2022. He has collaborated with the European Space Research and Technology Centre of the European Space Agency in the development of modal analysis tools for payload systems in satellites. In 2005, he held a postdoctoral research position involved in the area of new synthesis techniques in filter design with the University of Waterloo, Waterloo, ON, Canada. His research interests include applied electromagnetics, mathematical methods for electromagnetic theory, analytical and numerical methods for the analysis of microwave structures, and the design of waveguide components for space applications.

VICENTE E. BORIA (Fellow, IEEE) was born in Valencia, Spain, in May 1970. He received the Ingeniero de Telecomunicacin (Hons.) and Doctor Ingeniero de Telecomunicacin degrees from the Universidad Politècnica de València (UPV), Valencia, in 1993 and 1997, respectively. In 1993, he joined the Departamento de Comunicaciones, UPV, where he has been a Full Professor, since 2003. In 1995 and 1996, he was a Spanish Trainee with the European Space Research and Technology Centre, European Space Agency (ESTEC-ESA), Noordwijk, The Netherlands, where he was involved in the area of EM analysis and design of passive waveguide devices. He has authored or coauthored ten chapters in technical textbooks, 200 papers in refereed international technical journals, and more than 250 papers in international conference proceedings. His current research interests include the analysis and automated design of passive components, left-handed and periodic structures, and the simulation and measurement of power effects in passive waveguide systems. He has been a member of the IEEE Microwave Theory and Techniques Society (IEEE MTT-S) and the IEEE Antennas and Propagation Society (IEEE AP-S) since 1992. He is also a member of the European Microwave Association (EuMA) and the Technical Committees of the IEEE-MTT International Microwave Symposium and the European Microwave Conference. He serves as an Editorial Board Member of the *International Journal of RF and Microwave Computer-Aided Engineering*. He has been the Chair for the 48th European Microwave Conference held in Madrid, Spain. He acts as a regular reviewer for the most relevant IEEE and IET technical journals on the areas of interest. He was an Associate Editor of the IEEE MICROWAVE AND WIRELESS COMPONENTS LETTERS from 2013 to 2018 and the *IET of Electronics Letters* from 2015 to 2018. He serves as a Subject Editor for *Microwave Journal* and the *IET of Electronics Letters*.

MARCO GUGLIELMI (Life Fellow, IEEE) was born in Rome, Italy, in December 1954. He received the "Laurea in Ingegneria Elettronica" degree from the University of Rome "La Sapienza," Rome, in 1979, the M.S. degree in engineering from the University of Bridgeport, Bridgeport, CT, USA, in 1982, and the Ph.D. degree in electrophysics from Polytechnic University, Brooklyn, NY, USA, in 1986. He also attended the Scuola di Specializzazione in Elettromagnetismo Applicato, University of Rome "La Sapienza," in 1980. From 1984 to 1986, he was an Academic Associate with Polytechnic University, where he was an Assistant Professor, from 1986 to 1988. From 1988 to 1989, he was an Assistant Professor at the New Jersey Institute of Technology, Newark, NJ, USA. In 1989, he joined the RF System Division, European Space Agency, European Space Research and Technology Centre (ESTEC), Noordwijk, The Netherlands, as a Senior Microwave Engineer, where he was in charge of the development of microwave filters and electromagnetic simulation tools. In 2001, he was appointed as the Head of the Technology Strategy Section, ESTEC, where he contributed to the development of management processes and tools for the formulation of a European Strategy for Space Technology Research and Development. In 1981, he was awarded a Fulbright Scholarship in Rome and a Halsey International Scholarship Program (HISP) from the University of Bridgeport. In 2014, he retired from the European Space Agency and is currently an Invited Senior Researcher at the Universidad Politècnica de València (UPV), Valencia, Spain. He has been elevated to the grade of a fellow of the IEEE "For Contributions to Multimode Equivalent Networks and Microwave Filter Design," in January 2013.

Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom - August 03-05, 2023
Paper No. EEE 116
DOI: 10.11159/eee23.116

VDTA-C based voltage mode Tow-Thomas Biquad filter

Puneet Pandey, Dr.Bhavnesh Jaint, Dr.Garima Mann
Department of Electrical Engineering,
Delhi Technological University, New Delhi, India
erasepuneet@gmail.com; Bhaveshjaint@dtu.ac.in; Garimasolanki@dtu.ac.in

Abstract - This paper presents a new voltage mode Tow-Thomas biquad circuit. The proposed circuit employs three voltage differencing transconductance amplifier(VDTA) as active element together with two capacitors only as passive elements. This configuration provides independent tunability of centre frequency(w), quality factor(Q) and gain by properly setting the transconductance values of the VDTAs. The synthesis of the filter has been done using gpdk 180nm technology in Cadence virtuoso environment.

Keywords**:** VDTA, Current Conveyors, OTA, Biquad fiters, Tow-Thomas,Transconductance

1. Introduction

The TT biquad, known as the Tow-Tomas biquad, is a widely used filter implementation topology that has gained significant popularity [1]. Various active elements, including operational transconductance amplifiers (OTAs), current conveyors (CCs), and differential voltage current conveyors (DVCCs), have been employed in different realizations of this structure [1]. A recent development in this area involves the implementation of a new circuit using voltage differencing current conveyors (VDCCs) [2]. The field of microelectronics has witnessed notable advancements, leading to the introduction of novel circuit principles for active building blocks, aiming to enhance the performance of existing elements such as OTAs and CCs in fast analog signal processing. One of these recent advancements is the voltage differencing transconductance amplifier (VDTA) [3].

The VDTA consists of a current source controlled by the voltage difference between two input voltages and a multiple-output transconductance amplifier. It offers electronic tuning capabilities through its transconductance gains. In the VDTA, the differential input voltage (Vp, Vn) is converted to current at terminal Z by the first transconductance gain. The voltage drop at terminal Z is then converted to current at terminals X+ and X- (negative of X+) by the second transconductance gain. Both transconductance gains can be electronically controlled using external bias currents [5].

In this study, a voltage mode TT biquad filter is designed using VDTAs, incorporating an optimal number of active and passive elements. This circuit configuration enables high cut-off frequency and independent tunability.

2. Voltage Differencing Transconductance Amplifier (VDTA)

The circuit symbol for VDTA is depicted in Figure 1, illustrating input terminals labeled as p and n, and output terminals labeled as Z, X+, and X-. It is important to note that all these terminals demonstrate a high impedance characteristic.

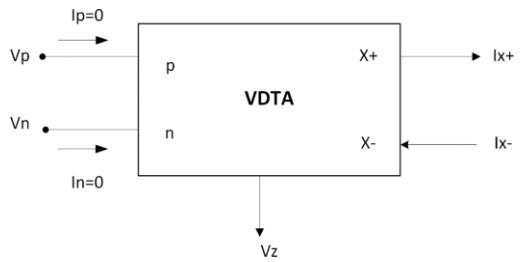


Fig. 1: VDTA block diagram

Considering ideal VDTA, the terminal relations can be characterized by the following matrix:

$$\begin{bmatrix} I_z \\ I_{x+} \\ I_{x-} \end{bmatrix} = \begin{bmatrix} g_{mf} & -g_{mf} & 0 \\ 0 & 0 & g_{ms} \\ 0 & 0 & -g_{ms} \end{bmatrix} \begin{bmatrix} V_p \\ V_n \\ V_z \end{bmatrix} \quad (1)$$

In the given matrix, the parameters g_mf and g_ms represent the tunable transconductance gains of the first and second stage respectively, and they are influenced by external currents supplied to the VDTA. The output current (Iz) is produced by converting the difference between the input voltages (Vp –Vn) using the transconductance gain g_mf. Different methods have been utilized for the realization of VDTA. Figure 2 [5] depicts the implementation of the VDTA.

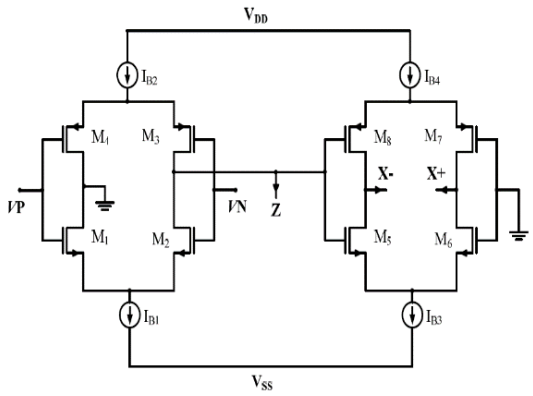


Fig. 2: VDTA Internal Architecture

The transconductance parameters derived through this structure are shown as:

$$g_{mf} = \left(\frac{g_1 g_2}{g_1 + g_2}\right) + \left(\frac{g_3 g_4}{g_3 + g_4}\right) \quad (2)$$

$$g_{ms} = \left(\frac{g_5 g_6}{g_5 + g_6}\right) + \left(\frac{g_7 g_8}{g_7 + g_8}\right) \quad (3)$$

where

$g_i = \sqrt{\frac{\mu C_{ox} W_i I_{Bi}}{L_i}}$ is the transconductance value.

In the equation above, I_Bi represents the bias current, μ denotes the effective carrier mobility, Cox corresponds to the gate-oxide capacitance, and W_i/L_i represents the effective channel length ratio of the i-th MOS transistor (where i = 1, 2, ..., 8).

2.1. VDTA-C Tow Thomas Biquad Filter

The figure provided in Fig. 3 illustrates the suggested arrangement of the multifunction voltage mode VDTA-C TT biquad filter. It enables the acquisition of both band-pass and low-pass filter responses concurrently. It may be noted that no resistors have been used in this configuration and the capacitors are grounded.

The analysis of the proposed circuit configuration yields the following output transfer function for band-pass filter:

$$T_{BP}(s) = \frac{V_{BP}}{V_{in}} = \frac{-\left(\frac{g_{mf1}}{C_1}\right)s}{s^2 + s\left(\frac{g_{ms1}}{C_1}\right) + \left(\frac{g_{mf3}\, g_{mf2}}{C_1 C_2}\right)} \quad (4)$$

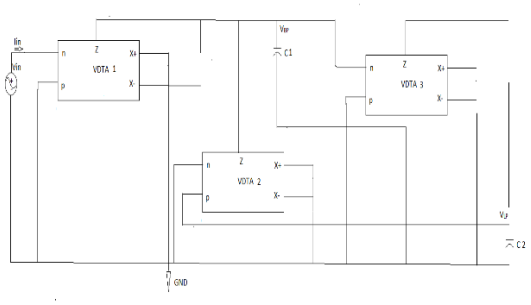


Fig. 3: VDTA-C based voltage mode Tow – Thomas biquad filter

Similarly, the transfer function of a low-pass filter can be derived as follows:

$$T_{LP}(s) = \frac{V_{LP}}{V_{in}} = \frac{\left(\frac{g_{mf1}\, g_{mf3}}{C_1 C_2}\right)}{s^2 + s\left(\frac{g_{ms1}}{C_1}\right) + \left(\frac{g_{mf3}\, g_{mf2}}{C_1 C_2}\right)} \quad (5)$$

The filter parameters namely the pole frequency (ω0), Bandwidth (BW) and quality factor (Q), can be mathematically expressed as follows:

$$\omega_0 = \sqrt{\frac{g_{mf3}\, g_{mf2}}{C_1 C_2}} \quad (6)$$

$$BW = \frac{g_{ms1}}{C_1} \quad (7)$$

$$Q = \sqrt{\frac{C_1\ g_{mf2}}{C_2 g_{mf3}}} \quad (8)$$

The DC gain for low pass response is given by:

$$H_{LP} = \frac{g_{mf1}}{g_{mf2}} \quad (9)$$

where g_mf1 and g_ms1 are first and second stage transconductance gain of VDTA1 and g_mf2 〖,g〗_mf3 are first stage transconductance gains of VDTA2 and VDTA3 block respectively.

In the case of a low-pass filter, it is possible to adjust the pole frequency and DC gain independently without any mutual interference, as demonstrated by equations (6), (7), and (9). Similarly, for a band-pass filter, we have observed the ability to independently tune the pole frequency and center frequency gain.

3. Simulation Results

The proposed circuit is simulated using gpdk 180 nm CMOS process in the Cadence virtuoso environment. The implementation of VDTA is done using given MOS transistor aspect ratio in Table 1. The supplied value of biasing is 150μA and at the supply voltage Vdd = +0.9V and Vss =-0.9V. The cutoff frequency is obtained as 10MHz using capacitor values as, C1 = C2 = 15pF. A sinusoidal

waveform of amplitude 1V and 100 KHZ is applied to the given circuit. The ideal Simulation results show that this choice yields transconductance values of VDTA as g_mf = g_ms= = 636.3 μA/V and the parasitic capacitance at the Z terminal is specified as Cp = 0.15 pF. The DC transfer characteristic is shown in the Fig. 4.

Table 1

MOS- Transistor Aspect Ratios (W/L) of VDTA

S.No	Transistors	L (*μm*)	W (*μm*)
1	M_1 - M_2 - M_5 - M_6	0.36	3.6
2	M_3 - M_4 - M_7 - M_8	0.36	16.64

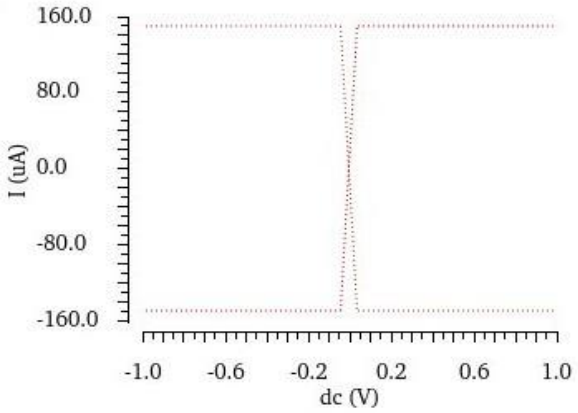


Fig. 4: DC characteristics of VDTA

The frequency response of band-pass and low pass filter is shown in Fig. 5. The phase response of band-pass filter is also obtained and can be seen in Fig. 6. The transient response of the filter is also obtained as shown in Fig. 7.

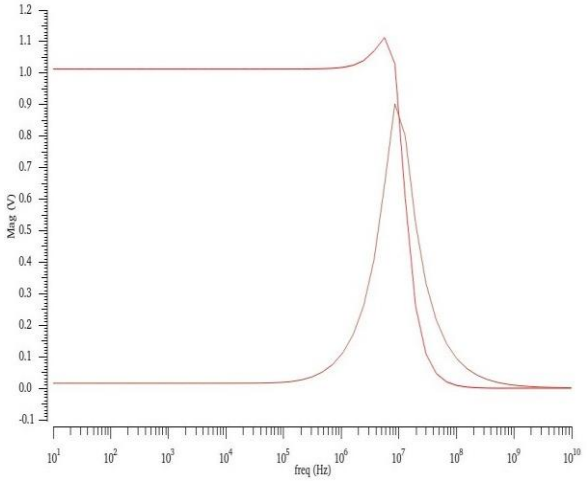


Fig. 5: Frequency response of TT Biquad filter

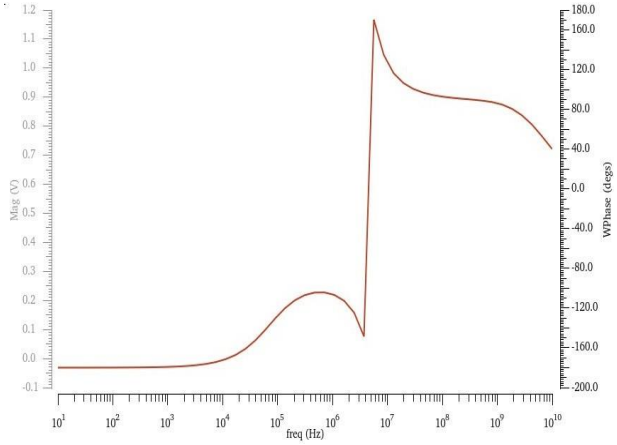


Fig. 6: Phase response of Band Pass Filter

Since the inverse configuration is obtained, we are aving 180 dergree phase change as we can see from the phase response and also from the transient response as seen in fig 8.

Table 2 - Comparision of TT biquad circuit based on number of components used

Active elements	No. of blocks	Resistors	Capacitor	Gain Control
OTA	3	6	2	Yes
OTRA	2	4	2	Yes
CCII	3	4	2	Yes
DVCC	3	3	2	No
VDCC	2	5	2	Yes
VDTA (Proposed)	3	0	2	Yes

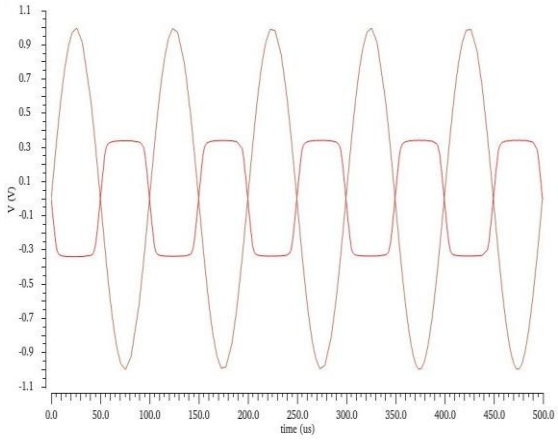


Fig. 7: Transient response of Band Pass Filter

4. Conclusion

The VDTA-C Tow-Thomas biquad presented in this paper makes use of an optimal combination of active and passive elements. Table 2 presents a comparison of various voltage mode TT biquad filter configurations, considering different numbers of active building blocks and passive elements.Noted point is that resistor is not used in this proposed configuration. The independent tunability is also demonstrated and the high frequency performance is required which can be seen in simulation results obtained from Cadence Virtuoso environment.

References

[1] Ahmed M. Soliman, "History and progress of the Tow-Thomas biquadratic filter part II: OTRA, CCII, and DVCC realizations", World Scientific, Journal of circuits, systems, and computers, Vol. 17, No.5, 2008.

[2] Dalibor Biolek, Raj Senani, Viera Biolkova, Zdenek Kolka,"Active elements for analog signal processing Classification, Review and New Proposals" Radioengineering, Vol.17, No.4, December 2008.

[3] Adel S. Sedra, Kenneth C. Smith, "Microelectronics circuits - theory and applications", Oxford University Press, 2013.

[4] Dinesh Prasad, Javed Ahmad, "New electronically-controllable lossless synthetic floating inductance circuit using single VDCC", Scientific Research, Circuits and Systems, 2014, 5, 13-17, January 2014.

[5] A. Yesil, F. Kacar, H. Kuntman, New simple CMOS realization of voltage differencing transconductance amplifier and its RF filter application. Radioengineering 20(3), 632–637 (2011)

[6] M. Srivastava, D. Prasad, Laxya and A. K. Sharma, "Novel Active Circuit for Realizing Variable Grounded Passive Elements," 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), 2016, pp. 559-563, doi: 10.1109/ICMETE.2016.110.

[7] D. Prasad et al., "Novel active PID controller employing VDTA," 2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2016, pp. 286-289, doi: 10.1109/ISSPIT.2016.7886050.

[9] K. Roja, M. Sarada and A. Srinivasulu, "A Voltage Mode All Pass Filter Employing Voltage Difference Transconductance Amplifier," 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 2018, pp. 1-4, doi: 10.1109/ECAI.2018.8679013.

[10] O. Channumsin and W. Tangsrirat, "Electronically Controllable SITO Multifunction Filter with Single VDTA," 2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST), 2018, pp. 1-4, doi: 10.1109/ICEAST.2018.8434433.

[11] P. Moonmuang, S. Unhavanich and W. Tangsrirat, "Simple Current-Controlled VDTA-R Schmitt Trigger Circuit," 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST), 2019, pp. 1-4, doi: 10.1109/ICEAST.2019.8802589.

[12] M. Faseehuddin, J. Sampe, S. Shireen and S. H. M. Ali, "Lossy and lossless inductance simulators and universal filters employing a new versatile active block", Informacije MIDEM, vol. 48, no. 2, pp. 97-113, 2018.

[13] "Grounded and Floating Inductance Simulation Circuits Using VDTA written by Dinesh Prasad, D. R. Bhaskar published by Circuits and Systems, Vol.3 No.4, 2012S,

[14] Tiwari and T. Singh Arora, "Dual-Mode Sinusoidal Oscillator Employing Single-Active Element and Grounded Capacitors," 2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES), 2021, pp. 1-5, doi: 10.1109/TRIBES52498.2021.9751652.

[15] T. Pukkalanun, P. Moonmuang and W. Tangsrirat, "Grounded Series/Parallel RL/RC Immittance Simulations with a Single VDTA," 2022 19th International Conference on Electrical Engineering/Electronics, Computer Telecommunications and Information Technology (ECTI-CON), 2022, pp. 1-4, doi: 10.1109/ECTI-CON54298.2022.9795367.

AI in Civil Engineering

ORIGINAL ARTICLE

Open Access

Water surface profile prediction in non-prismatic compound channel using support vector machine (SVM)

Vijay Kaushik[1*] and Munendra Kumar[1]

Abstract

The process of estimating the level of water surface in two-stage waterways is a crucial aspect in the design of flood control and diversion structures. Human activities carried out along the course of rivers, such as agricultural and construction operation, have the potential to modify the geometry of floodplains, leading to the formation of compound channels with non-prismatic floodplains, thus possibly exhibiting convergent, divergent, or skewed characteristics. In the current investigation, the Support Vector Machine (SVM) technique is employed to approximate the water surface profile of compound channels featuring narrowing floodplains. Some models are constructed by utilizing significant experimental data obtained from both contemporary and previous investigations. Water surface profiles in these channels can be estimated through the utilization of non-dimensional geometric and flow parameters, including: converging angle, width ratio, relative depth, aspect ratio, relative distance, and bed slope. The results of this study indicate that the SVM-generated water surface profile exhibits a high degree of concordance with both the empirical data and the findings from previous research, as evidenced by its R^2 value of 0.99, RMSE value of 0.0199, and MAPE value of 1.263. The findings of this study based on statistical analysis demonstrate that the SVM model developed is dependable and suitable for applications in this particular domain, exhibiting superior performance in forecasting water surface profiles.

Keywords Non-prismatic compound channel, Non-dimensional parameter, Support vector machine (SVM), Water surface profile

1 Introduction

With growing population across the world, increasingly more people are living in and exploring the areas along rivers, leading to serious problems related to natural flooding, which often results in heavy life losses, in addition to economic damage. Trend analyses have shown that the proportion of both global losses and deaths attributable to flood disasters continue to rise rapidly (Berz, 2000). Given that flooding occurs when the volume of water flowing through a channel is greater than the channel's capacity, predicting the flow rate of natural streams is crucial for flood prevention. As a means to reduce casualties and salvage assets as much as possible, accurate flow parameter prediction in flood circumstances, therefore, has aroused the attention of academics and engineers over recent years, with various techniques and methodologies devised to precisely gauge and anticipate river discharge, velocity distribution, shear stress distribution, and water surface level in the event of overbank floods. Compound channels are a prevalent river feature during the period of overbank flows. The hydrodynamics of a river's discharge can induce alterations to the morphology of floodplains, leading to the development of a two-stage channel system that

*Correspondence:
Vijay Kaushik
vijaykaushik_2k20phdce01@dtu.ac.in
[1] Department of Civil Engineering, Delhi Technological University, Delhi 110042, India

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

exhibits either converging or diverging characteristics. The transfer of energy from main channels to flood zones is comparatively higher in a composite waterway with narrowing flood zones, and thereby rendering the replication of flows in such a channel is a challenging task. Although the flow models of straight and meandering prismatic two-stage channels have been investigated by some researchers, including Sellin (1964), Myers and Elsawy (1975), Knight et al. (2010), and Khatua et al. (2012), very little is known about non-prismatic compound channels. When a channel is shaped like a converging or diverging path, the floodplain flow would increase and the floodplain flow would decrease (James & Brown, 1977). After investigating compound channels with symmetrically descending floodplains, Bousmar and Zech (2002), Bousmar et al. (2004), Rezaei (2006), and Rezaei and Knight (2009) discovered additional losses of head as well as the transfer of momentum from main channels to floodplains. Proust et al., (2006) have surveyed asymmetric geometry at a faster convergence rate, finding that there were more mass transfers and more head losses at a large convergence angle of 22 degrees. Flows in curved, two-stage converging and diverging channels were investigated by Chlebek et al. (2010), who found that when floodplains expand, the velocity gradient would rise and the head losses would increase, owing to mass and momentum transfer. There are discernible variations in the flow distribution between main channels and floodplains as a result of varying flow forces in each segment. Rezaei and Knight (2011) have explored the depth-averaged velocity, local velocity distributions, and boundary shear stress distributions in compound channels with nonprismatic floodplains in a range of convergence degrees, discovering an increase in velocity near the main channel walls, especially during the second half of the convergence reach; in addition, the effects of the lateral flows that enter main channels were observed at a high relative depth, as shown with the depth-averaged velocities. Overbank flows in compound channels with non-prismatic floodplains were studied by Yonesi et al. (2013), who found that when the depth ratio is increased or the roughness ratio is decreased, the velocity gradient between main channels and floodplains may be reduced at the divergence stretch's middle and end sections. As the divergence angle becomes greater, so does the velocity gradient. The proportion of the split discharge is affected by the roughness ratio and the depth ratio. Furthermore, the shear stress gradient would increase with higher roughness of floodplains. Naik and Khatua (2016) have developed a multivariable regression model to forecast the water surface profile for compound channels with non-prismatic floodplains. Their proposed model matches actual facts and other researchers' conclusions well. In non-prismatic compound channels, high-flow features have received little attention, and more data are needed for prediction of water surface profiles. To recreate precise profiles, more tests shall be conducted on non-prismatic compound channels with converging floodplains.

Building profile models for water surface by using different approaches makes it difficult to analyze dependencies among components, and these models would get complicated and time-consuming. This cuts in half the time needed for experiments and labour-intensive calculations. Support vector machine (SVM), gene expression programming (GEP), artificial neural networks (ANN), fuzzy neural networks (ANFIS), and M5 tree decision models are all often used to estimate flows in composite channels (Parsaie et al., 2017b). Unlike other soft computing algorithms like ANN and SVM, GEP can really produce mathematical relationships (Karimi et al., 2015), although GEP-based river engineering has been neglected (Azamathulla et al., 2013; Pradhan & Khatua, 2017). Parsaie et al. (2015) have used the support vector machine (SVM) method to forecast discharges in a compound open channel, finding that SVM can deliver the maximum accuracy, as illustrated by the error indices. By employing various analytical techniques, including the single channel method (SCM), coherence method (COHM), and divided channel method (DCM), Parsaie et al. (2016) have forecasted the discharges of a compound open channel. To enhance the precision of the flow discharge prediction, they have also devised a radian basis neural network (RBF). Finally, they found that the DCM featuring a horizontally separated boundary between subsections exhibited a correlation of determination (R^2) of 0.76, with more precise results acquired in analytical methods. The evaluation of the MLP model's outcomes exhibited a higher degree of precision ($R^2 = 0.95$) than the RBF ($R^2 = 0.85$) and other analytical methodologies. In another study, Parsaie and Haghiabi (2017) employed certain analytical techniques, specifically artificial neural network (ANN) and multivariate adaptive regression splines (MARS), to predict the discharge capacity of compound open channels, finding that the divided channel method yielded a coefficient of determination (R^2) value of 0.76 and a root mean square error (RMSE) value of 0.162, exhibiting superior performance over other analytical methods evaluated. Their research further revealed that the precision of the applied soft computing models surpassed that of analytical methods by delivering an R^2 value of 0.97 and an RMSE value of 0.03, indicating that the MARS model has a higher level of accuracy than the MLP model. In an exploration conducted by Parsaie et al. (2017a) to evaluate the precision of various empirical techniques, including the Single

Channel Method (SCM), Divided Channel Method (DCM), and Coherence Method (CM), in approximating the flow discharges in a heterogeneous compound channel roughness, they found that if assuming the virtual perpendicular line as the boundary line between sub-sections, DCM could result in a coefficient of determination greater than 0.90, demonstrating its effective performance through empirical approaches. By using the most important characteristics, such as the width ratio, relative flow depth, aspect ratio, Reynolds number, and Froude number, Khuntia et al. (2018) created an artificial neural network (ANN) model to forecast boundary shear stress distribution in straight compound channels, revealing that back-propagation neural network (BPNN) models could function effectively over wide parameter spaces. Das and Khatua (2018b) designed a numerical model that relies on the momentum balance principle to forecast water surface elevations of compound channels with the converging floodplains. The numerical model was solved with the finite difference method in MATLAB, and the simulation results demonstrated a high level of concurrence with the experimental datasets. In addition, the model could yield the lowest level of errors in terms of such metrics as the mean absolute error, mean absolute percentage error, and root mean square error. To calculate flow magnitudes and velocities in the upper and lower main channels, Das et al. (2018) have enhanced the conventional independent subsection method (ISM) framework, which features the establishment of four distinct subsections that are delineated by both vertical and horizontal division lines. These lines correspond to the vertical interface between the primary channel and floodplains, as well as the horizontal interface between upper and lower main channels. The computation results demonstrate that the method possesses the ability to accurately forecast the discharge distributions in both floodplains and main channels. Das and Khatua (2018a) have conducted experiments on symmetric compound channels in diverging and converging floodplains with various relative flow depths, in order to examine the resistance characteristics of overbank flows in non-prismatic sections. A multivariable regression model was constructed, with consideration given to the geometric and hydraulic parameters previously mentioned, for the purpose of estimating the Manning's roughness coefficient in non-prismatic compound channels. This model demonstrated favourable outcomes in comparison to the alternative methodologies across diverse experimental channels and field data, exhibiting reduced errors. Das et al. (2019) have developed a formula for calculating the flow rate in a system of compound channels with both divergent and convergent branches. Das et al. (2020) employed artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) techniques to forecast the discharge in compound channels featuring converging and diverging sections. The discharge was influenced by several significant input parameters, including the friction factor ratio, hydraulic radius ratio, relative flow depth, and bed slope. The ANFIS model demonstrated a favourable level of performance in comparison to the ANN model, as evidenced by a high coefficient of determination of 0.86 and a low root mean square error of 0.005. In order to estimate the water surface profile in converging compound channels by using non-dimensional variables, Naik et al. (2022) suggested a unique equation based on GEP. Yonesi et al., (2022) have modelled the discharges of compound open channels with non-prismatic floodplains (CCNPF) by utilizing some soft computation models, such as multivariate adaptive regression splines (MARS) and group method of data handling (GMDH). The outcomes were subsequently compared with the data from multilayer perceptron neural networks (MLPNN). The statistical error indices indicate that the developed models can deliver sufficient accuracy in estimating the flow discharge in CCNPF. Additionally, an analysis of the architecture of advanced GMDH and MARS models revealed that the following parameters exerted the most significant influence on the modelling and estimation of discharges: roughness, area, hydraulic radius, flow aspect ratio, depth, and angle of convergence or divergence of floodplain. In a study conducted by Bijanvand et al. (2023), soft computing models such as the Multi-Layer Perceptron Neural Network (MLPNN), Group Method of Data Handling (GMDH), Neuro Fuzzy Group Method of Data Handling (NF-GMDH) and Support Vector Machine (SVM) were utilised to predict water surface elevation of compound channels with converging and diverging floodplains. In this research, the SVM model exhibited superior performance in the testing phase, as evidenced by the statistical indicators; the utilization of the adaptive fuzzy methodology in the development of the Group Method of Data Handling (GMDH) model resulted in a noteworthy enhancement of precision, as evidenced by the statistical metrics during the testing phase; The optimal performance of the activation and kernel functions in the construction of the MLPNN model and the SVM was found to be associated with the sigmoid and radial tangent functions. In their study, Kaushik and Kumar (2023a) utilized machine learning techniques to forecast the water surface profile of a compound channel with a converging floodplains based on a combination of geometric and flow variables. The results indicate that the proposed artificial neural network (ANN) model could deliver a high degree of accuracy in forecasting the water surface profile, as compared to gene expression

programming (GEP), support vector machine (SVM), and previously established techniques. Moreover, Kaushik and Kumar (2023b) have also employed gene expression programming to formulate a novel equation for compound channels with converging floodplains. Expressed in terms of nondimensional geometric and flow variables, the equation can be used to estimate the boundary shear force conveyed by floodplains. Finally, it is found that the anticipated proportion of the floodplain-supported shear force as identified through GEP exhibits a satisfactory level of concurrence with the empirical evidence.

This study will employ Support Vector Machine (SVM) modelling technique to predict the water surface profile of non-prismatic compound channels with converging floodplains. This study, along with other researches on similar channel configurations, utilizes experimental datasets to investigate the impacts of diverse non-dimensional geometric and flow parameters on water surface elevations. Other different methods were previously utilized by other researchers to compute the water surface profile in non-prismatic compound channels; and they will be subjected to effectiveness or error analysis to assess the accuracy of the model proposed in this study.

2 Materials and methodologies

2.1 Data sources

Numerous explorations have been conducted on the distribution of shear force, transfer of momentum, and discharge mechanisms of compound channels with non-prismatic floodplains. This study will utilize the data from Rezaei (2006) and Naik and Khatua (2016), as well as the experiments of this paper's author. Table 1 showcases the data pertaining to the experimental channel dimensions of composite waterway with diverse geometric characteristics.

2.2 The experimental setup

The investigations for this study were conducted at a hydraulic laboratory of the Department of Civil Engineering at Delhi Technological University, India, with the experimental procedures operated in a masonry flume 12 m long, 1 m wide, and 0.8 m deep. A cross-sectional compound configuration was established with brick masonry in the flume, and the main channel was 0.5 m in width and 0.25 m in depth. The section of the channel to exhibit convergence was constructed with brick masonry. The maintenance of the geometry was achieved through the utilization of a converging angle of 4° in the channel. The compound channel under investigation exhibited a non-prismatic geometry, with a converging length of 3.6 m. The state of subcritical flow was measured under various circumstances within the two-stage channel, which featured a longitudinal bed slope of 0.001. Manning's n value was estimated based on the data obtained from both in-bank and over-bank flows in the flooded areas and primary channel. The water supply was recycled through a mechanism that pumps

Table 1 Channel dimensions used (Kaushik & Kumar, 2023a)

Verified test channel	Type of channel	Angle of convergent (θ)	Longitudinal slope (S)	Cross-sectional geometry	Total channel width (B) (m)	Main channel width (b) (m)	Main channel depth (h) (m)	Converging length (L) (m)	Aspect ratio (δ)
Channel 1 Rezaei (2006)	Converging	11.31°	0.002	Rectangular	1.2	0.398	0.05	2	7.96
Channel 2 Rezaei (2006)	Converging	3.81°	0.002	Rectangular	1.2	0.398	0.05	6	7.96
Channel 3 Rezaei (2006)	Converging	1.91°	0.002	Rectangular	1.2	0.398	0.05	6	7.96
Channel 1 Naik and Khatua (2016)	Converging	5°	0.0011	Rectangular	0.9	0.5	0.1	2.28	5
Channel 2 Naik and Khatua (2016)	Converging	9°	0.0011	Rectangular	0.9	0.5	0.1	1.26	5
Channel 3 Naik and Khatua (2016)	Converging	12.38°	0.0011	Rectangular	0.9	0.5	0.1	0.84	5
Present channel	Converging	4°	0.001	Rectangular	1.0	0.5	0.25	3.60	2

water from an underground sump to a reservoir located in the experimental channel. The flow originating from the conduit was accumulated within a volumetric receptacle equipped with a V-shaped notch; subsequently, it returned to the sump situated below. Figure 1 presents an exposition of the geometric characteristics and a sectional view of the composite waterway.

The experimental setup designed for the proposed research is depicted in Fig. 2 for its top view. A tailgate was implemented at the terminal point of the flume to regulate the profile of water surface and enforce a predetermined flow depth within the flume segment. The height of the water surface was determined with a point gauge, which could achieve an accuracy of 0.1 mm. An Acoustic Doppler Velocimeter (ADV) was employed to identify velocity distributions in three dimensions. The data obtained via the Acoustic Doppler Velocimeter (ADV) was subject to filtration in the Horizon ADV software. The lateral distributions of boundary shear stress were measured with a Preston tube of an outer diameter of 5 mm at the identical sections where the velocity distributions were examined. The pressure difference

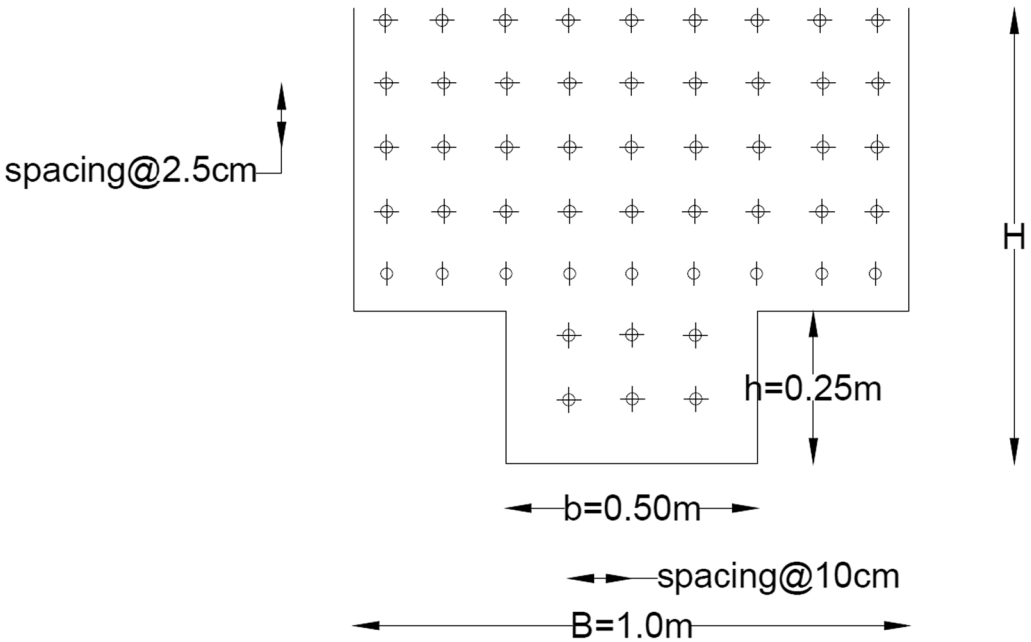


Fig. 1 Cross section of the compound channel

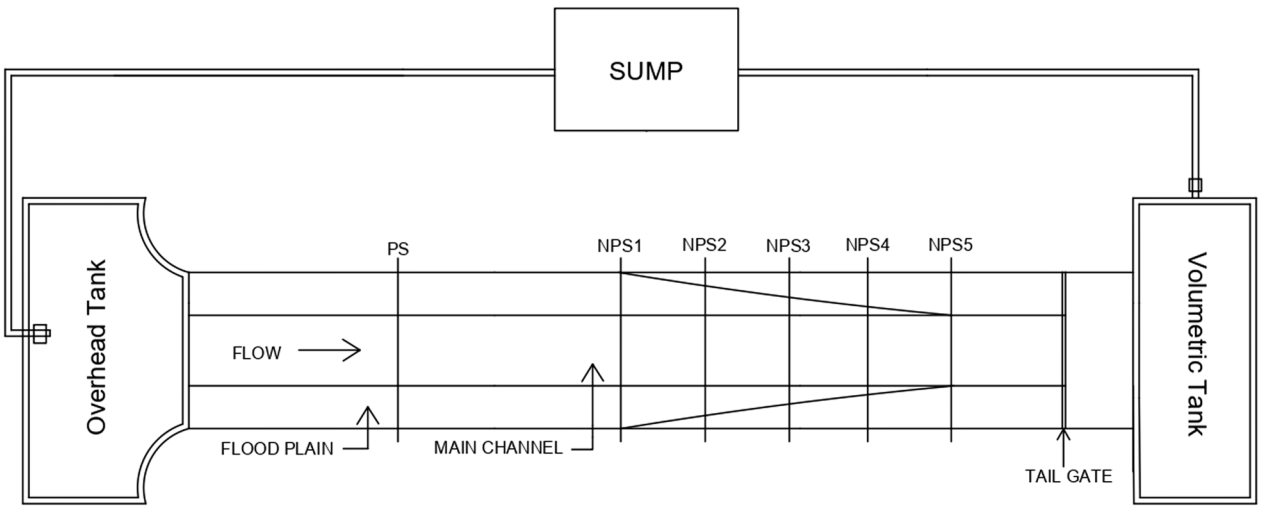


Fig. 2 Plan view of the experimental setup

was measured with a digital manometer. The utilization of Patel (1965) calibration equations was subsequently employed to compute the shear stress values. Measurements of local boundary shear stress and velocity were conducted along the wetted perimeter at intervals of 2.5 cm and 10 cm in the vertical and horizontal directions, respectively, as depicted in Fig. 1.

2.3 Theoretical background

For river specialists, the ability to make reliable projections of the contour of the top of the water during flood flows is essential for constructing drainage canals, flood defence systems, river training works, and floodplain management. As flood flow conditions in natural rivers are unpredictable, it may be difficult to acquire accurate and complete field data. Understanding the water surface profile in composite channels, especially those including both prismatic and non-prismatic floodplains, requires extensive laboratory experiments. Naik and Khatua (2016) have utilized multivariate analyses based on Eqs. (1–21) for forecasting the water surface profile in a converging compound channel, and Naik et al. (2022) have used the GEP approach based on Eq. (22).

$$\Psi = 1.06\alpha^{0.22} \tag{1}$$

$$\Psi = 1.16\alpha^{0.22} \tag{2}$$

$$\Psi = 1.21\alpha^{0.29} \tag{3}$$

$$\Psi = 0.07\alpha + 1.78 \tag{4}$$

$$\Psi = 0.05\alpha + 1.28 \tag{5}$$

$$\Psi = 0.13\alpha + 1.25 \tag{6}$$

$$\Psi = -0.14X_r + 1.22 \tag{7}$$

$$\Psi = -0.15X_r + 1.32 \tag{8}$$

$$\Psi = -0.22X_r + 1.37 \tag{9}$$

$$\Psi = -0.15X_r + 2.01 \tag{10}$$

$$\Psi = -0.16X_r + 1.45 \tag{11}$$

$$\Psi = -0.21X_r + 1.67 \tag{12}$$

$$\Psi = -1.22 + 2.27\alpha^{0.22} + 0.18X_r \tag{13}$$

$$\Psi = -1.21 + 2.28\alpha^{0.22} + 0.19X_r \tag{14}$$

$$\Psi = -0.58 + 1.63\alpha^{0.29} + 0.18X_r \tag{15}$$

$$\Psi = -0.66 + 0.29\alpha + 0.12X_r \tag{16}$$

$$\Psi = 0.86 + 0.29\alpha + 0.11X_r \tag{17}$$

$$\Psi = 0.86 + 0.29\alpha + 0.12X_r \tag{18}$$

$$\Psi^*(\theta) = \frac{Actual\Psi}{Eq.(12)} \tag{19}$$

$$\frac{Actual\Psi}{Eq.(12)} = e^{0.0017\theta} \tag{20}$$

$$\Psi = e^{0.0017\theta}\left(-1.21 + 2.25\alpha^{0.22} + 0.18X_r\right) \tag{21}$$

$$\Psi = \left[\frac{3.72 + \delta\beta + S\beta}{3.72 - S - 0.12\alpha} + \frac{9.63 - \alpha\delta}{(5.73 + \theta) \times \left(\delta^2\beta + 2\delta X_r\right)} - 14.75\delta S - \alpha\delta S\right] \tag{22}$$

A predictive model was formulated to estimate the configuration of the water's free surface in a composite waterway in convergent flooded areas. As anticipated, the fluid flow exhibits non-uniform behaviour within the non-prismatic region prior to its entry into the straight zone. The surface of both the primary waterway and the flooded zones showcases a consistently smooth texture. The mean value of Manning's n for the aforementioned smooth surfaces is 0.013. When assessing the water surface configuration of composite waterways in the narrowing floodplains, several significant factors shall be considered, including but not limited to: the width ratio (α), relative depth ratio (β), aspect ratio (δ), converging angle (θ), relative distance (X_r), and longitudinal bed slope (S_o).

The requisite equation in a dimensionless form can be formulated as follows:

$$\Psi = F(\alpha, \beta, \delta, \theta, X_r, S_o) \tag{23}$$

2.4 SVM model

The Support Vector Regression (SVR) algorithm is a commonly employed machine learning technique for the purpose of regression analysis. The method in question operates on the principles analogous to those of support vector machines (SVMs) for the classification purpose, albeit with the objective of forecasting a continuous output variable on the basis of input variables. The fundamental concept underlying Support Vector Regression (SVR) is to identify a function that can effectively map input variables to output variables, with minimal errors incurred, while also satisfying a predetermined margin. The process of attaining this objective involves the conversion of the input variables into a space of higher dimensions via a kernel function. Subsequently, the regression line that optimizes the margin while maintaining the error under a predetermined tolerance level is determined (Parsaie et al., 2015). SVR exhibits a notable benefit in its capacity to effectively manage non-linear associations between input and output variables. This feature is facilitated by such kernel functions as polynomial, radial basis function (RBF), and sigmoid function, which enable the capture of intricate relationships between variables. Support Vector Regression (SVR) has been effectively utilized in diverse domains, including finance, transportation, and medical imaging. It is of utmost significance to meticulously choose and adjust the kernel functions and other associated parameters, in order to attain optimal performance. One common method for transforming a linear classifier into a non-linear classifier utilizes a non-linear function to map the input space x into a feature space F. An alternative approach involves the utilization of a non-linear function for mapping. The separating function in space F can be expressed as below (Parsaie et al., 2015):

$$f(x) = w^T \varphi(x) + b \tag{24}$$

Within the subspace, the statistical methodology denoted by the algebraic function f (x, w) can be described as follows:

$$w = \sum_{i=1}^{n} \alpha_i x_i \tag{25}$$

$$f(x, w) = \sum_{j=1}^{n} \alpha_i x_i \varphi_i(x) + b \tag{26}$$

$$f(x) = \sum_{i=1}^{n} \alpha_i x_i^T x + b \tag{27}$$

In the feature space F, these expressions take the following forms:

$$f(x) = \sum_{i=1}^{n} \alpha_i \varphi(x_i)^T \varphi(x) + b 0 \le \alpha_i \le C \tag{28}$$

$$k(x, x\prime) = \varphi(x)^T \varphi(x\prime) \tag{29}$$

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x, x_i) + b \tag{30}$$

This study finds that among the numerous kernel functions present in SVM, the radial basis kernel function is the most efficacious. Conversely, alternative kernel functions exist that can be utilized to provide assistance in a broad sense.

Linear kernel: $k\left(x_i, x_j\right) = x_i^T x_j$
Polynomial kernel: $k\left(x_i, x_j\right) = \left(\gamma x_i^T x_j + r\right)^d, \gamma > 0$
RBF kernel: $k\left(x_i, x_j\right) = \exp(-\gamma \left\| x_i - x_j \right\|)^2, \gamma > 0$
Sigmoid kernel: $k\left(x_i, x_j\right) = \tan h(\gamma x_i^T x_j + r), \gamma > 0$

The SVM model employs parameters C, γ, r, and d. The efficacy of the Support Vector Machine (SVM) model is contingent upon the calibre of the selected settings for the said parameters, in addition to the kernel parameters. The selection of values for parameters C, γ, and r plays a crucial role in regulating the complexity of the regression model. However, the process of determining the optimal parameters would be further complicated by the intricate nature of their selection. The kernel functions are utilized to reduce the dimensionality of the input space and facilitate classification. The Support Vector Machine (SVM) model is further refined through its exposure to data sets, a process that is analogous to the refinement of other neural network models. This study utilized a reliable and high-quality experimental dataset to evaluate the water surface profile of composite waterways with convergent floodplains. Prior to building the SVM model, the data were partitioned into a training set and a testing set at a 70:30 ratio in MATLAB R (2019).

2.5 Statistical measures

To further test the accuracy of the developed SVM model, various metrics of error analysis, such as the coefficient of determination (R^2), mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE), are measured by using the following equations (Kaushik & Kumar, 2023a, 2023b):

$$R^2 = \frac{\sum_{i=1}^{N} (a_i - \overline{a})^2 (p_i - \overline{p})^2}{\sum_{i=1}^{N} (a_i - \overline{a})^2 \sum_{i=1}^{N} (p_i - \overline{p})^2} \tag{31}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|p_i - a_i| \tag{32}$$

$$\mathrm{MAPE}(\%) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|p_i - a_i|}{a_i} \times 100\right) \tag{33}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - a_i)^2} \tag{34}$$

The above equations utilize the actual and predicted values, which are denoted by a and p, respectively. Additionally, the means of the actual and predicted values, represented by $\bar{a}$ and $\bar{p}$, respectively, are also utilized in the computation. The total number of datasets is denoted by N.

3 Results and discussion

In Fig. 3, the surface profile of water in a non-prismatic composite channel is depicted for various relative flow depths in conjunction with the channel length. In the straight segment of the waterway, the water level remains constant. However, in the convergent section, particularly in the latter half of the transition, the velocity of the flow would become faster, and the water level would be decreased. The flow in the lower section of the waterway exhibits a degree of consistency, albeit with some variations. Figure 4 illustrates the variation of the non-dimensional water surface profile (Ψ) in relation to the width ratio, relative depth, aspect ratio, and relative distance at a convergence angle of $\theta = 4°$ and relative flow depths ranging from $\beta = 0.10$ to 0.60. An increase in flow depth would result in a wider width ratio, which would in turn broaden the water surface profile. The water surface always changes in the same way as the width ratio, even at extreme convergence angles. For a wide range of convergent angles, the water surface would rise in a way not linearly connected to the depths at which the two flows are interacting with each other. The water surface profile across the non-prismatic compound channel exhibits a decreasing aspect ratio as the water traverses various sections of the channel. The influence of distance on two-dimensional representations of the water surface would diminish with increased separations between the spots. This demonstrates that the velocity head would quickly rise up as the potential head is lowered and the transition converges. Given sharper angles of convergence in floodplains, the decline is very dramatic.

Figure 5 shows a scatter plot to compare the predicted and actual values for the SVM model. The close proximity of the values to the line verifies an excellent agreement, remarkably indicating the SVM model's ability to

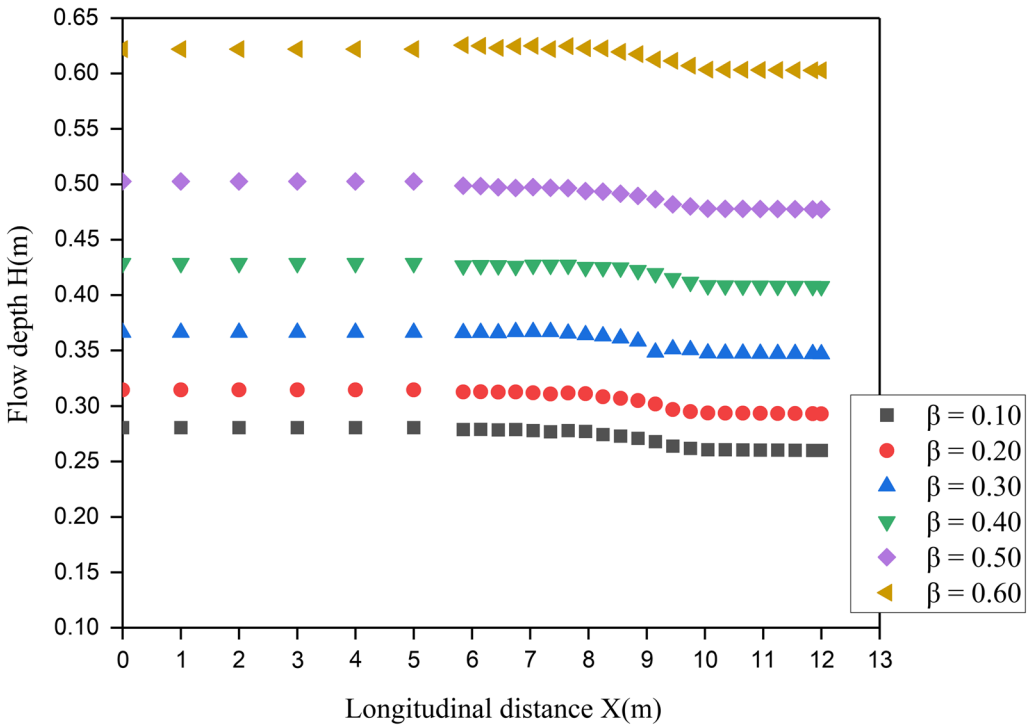


Fig. 3 Water surface profile of the two-stage waterway

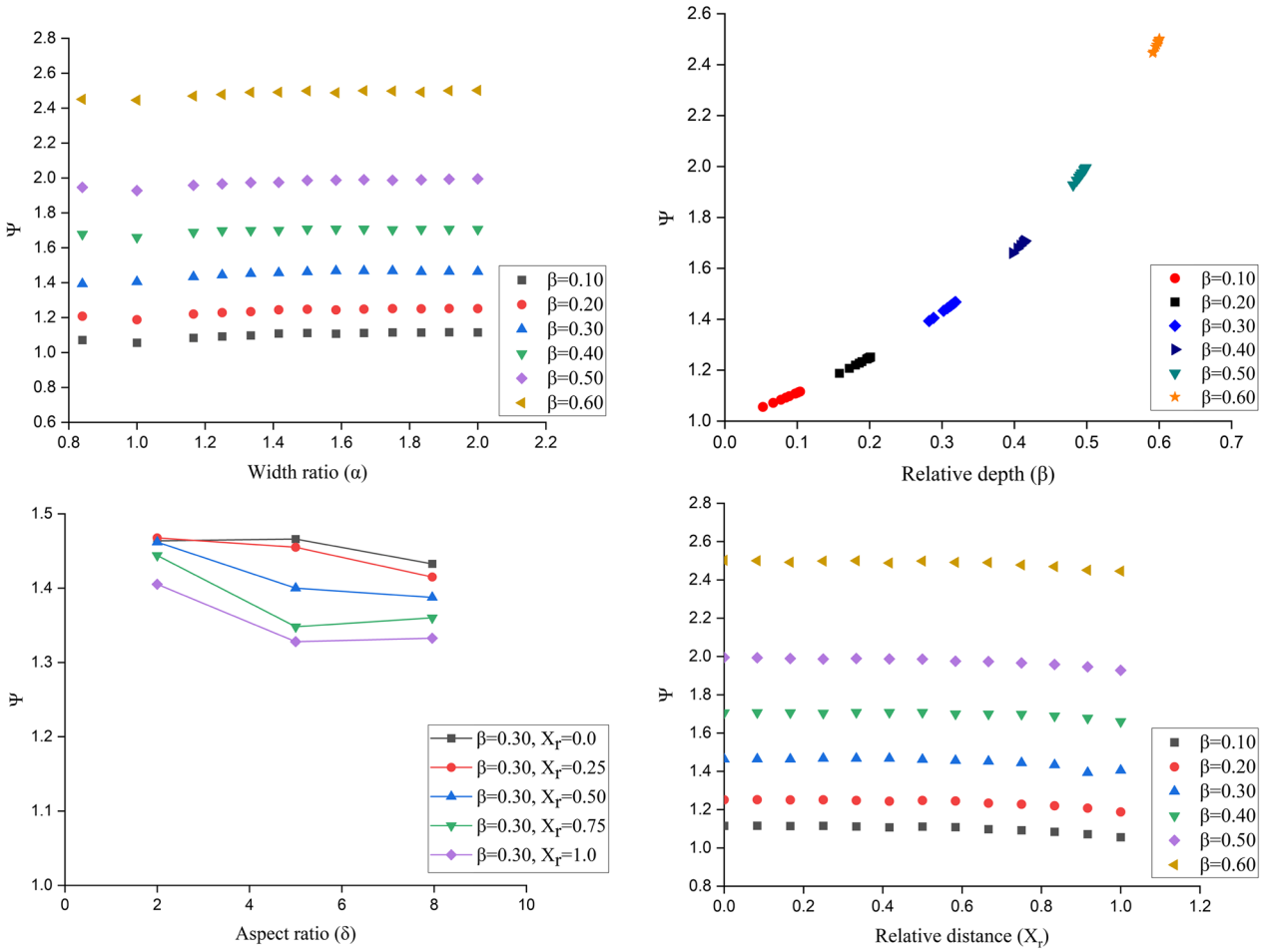


Fig. 4 Variation of non-dimensional water surface profile with various non-dimensional parameters

deliver accurate predictions. The fact that this model may offer analytical, non-parametric equations that can be used in real-world situations makes it much more useful. When making forecasts, the value of the SVM approach would yield more promising results, since the projected values are extremely near to the best-fitting line. Table 2 showcases the statistical metrics used to assess the performance of the SVM model, including R^2, MSE, RMSE, MAE, and MAPE. As shown in the table, SVM ($R^2 = 0.99$ and RMSE = 0.0199) predicts a better water surface profile in non-prismatic compound channels than other approaches, such as those proposed by Naik et al. (2022) and Naik and Khatua (2016). SVM is found to be the most suitable approach for predicting the surface water profile of composite waterways with the narrowing floodplains, with a mean absolute percentage error (MAPE) of 1.263, which is better that that of other methods.

4 Conclusion

This study demonstrates the applicability of Support Vector Machine (SVM), a machine learning technique, in detecting the elevation of water surface of two-stage channels with the narrowing floodplains. The proposed models are derived from high-quality laboratory datasets, which feature dimensionless geometric and flow characteristics and pertain to composite waterways that exhibit diverse convergent angles and relative flow depths. Several findings and implications are drawn from this study.

Certain parameters, such as the width ratio, relative flow depth, aspect ratio, angle of convergent, relative longitudinal distance, and longitudinal bed slope, appear to influence the proposed model. The phenomenon of decreasing flow depths along the channel length is attributable to the converged geometry of the channel. In fact, this trend has been observed across various floodplain

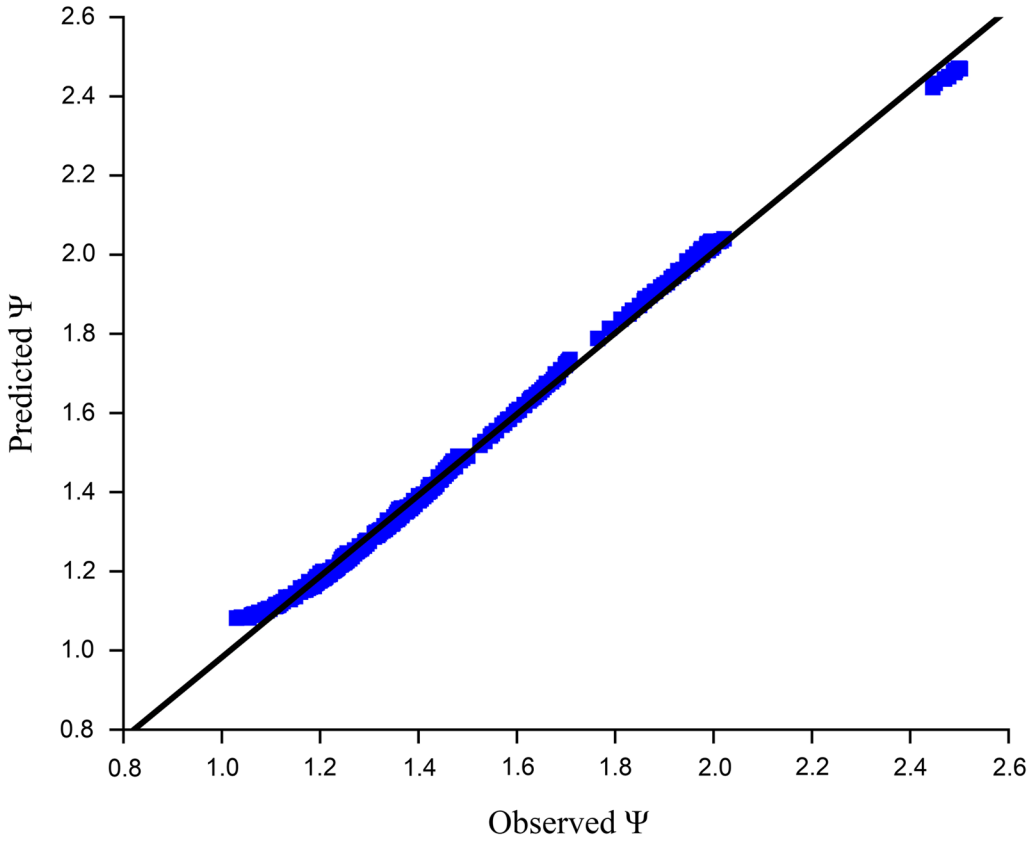


Fig. 5 Comparison between the predicted Ψ and the observed Ψ

convergence angles and at higher relative depths. It is also revealed that the non-dimensional profile of water surface exhibits an upward trend with hiking width ratio and relative depth, while showing a downward trend in terms of the aspect ratio and relative distance. The alteration of water surface's elevation in two-stage waterways exhibits the similar trend across the spectrum of narrowing angles, as demonstrated in other previous investigations. By exploring the correlation between the elevation of water surface and the non-dimensional geometric and hydraulic features of a narrowing composite waterway, this study identified non-linear correlations among the parameters. The models generated in this study deliver superior performance in terms of R^2, MAE, RMSE, and MAPE metrics over previous methodologies, such as those proposed by Naik and Khatua (2016) and Naik et al. (2022). These findings validate that the Support Vector Machine (SVM) approach can yield a satisfactory estimation of the water surface profile in composite waterways, as determined by the assessment metrics. The SVM model proposed in this study exhibits superior performance over other models, as evidenced by its R value of 0.99, RMSE value of 0.0199, and MAPE value of 1.263. These results prove the models' satisfactory performance and practical usefulness across the range of non-dimensional factors examined in this study. The utilization of statistical regression as the sole methodology is considered obsolete and should be improved. The aforementioned process is operated with a technique that could translate knowledge into a feature space of high dimensionality. Subsequently, various methods can be employed to identify interrelated fragments of information within this framework. The comprehensive mapping results in categorized linkages. The physical,

Table 2 Error analysis of SVM-predicted Ψ against previously developed methods

Statistical parameters	SVM model	Naik et al. Method (2022)		Naik and Khatua Method (2016)
		Training	Testing	
R^2	0.99	0.99	0.99	0.896
MSE	0.00039	0.0008	0.0007	0.0019
RMSE	0.0199	0.028	0.027	0.043
MAE	0.017	0.022	0.022	0.0015
MAPE (%)	1.263	1.543	1.546	2.429

process-oriented research in riverine hydraulics has predicted that various parameters have an impact on energy losses, velocity, and discharge. The challenges in obtaining reliable hydraulic data in riverine environments pose significant obstacles to in-depth research in this field.

Acknowledgements
The authors express their gratitude for the assistance provided by the Department of Civil Engineering at Delhi Technological University, located in Delhi, India.

Author contributions
The analysis was conducted and the manuscript was prepared by VK. The manuscript underwent a thorough review by MK, who granted final approval for its submission.

Funding
Not applicable.

Availability of data and materials
All relevant data are included in the manuscript or its additional information.

Declarations

Competing interests
The authors reported no potential conflicts of interest.

Received: 25 March 2023 Revised: 10 June 2023 Accepted: 1 July 2023
Published online: 28 July 2023

References

Azamathulla, H. M., Ahmad, Z., & Ghani, A. A. (2013). An expert system for predicting Manning's roughness coefficient in open channels by using gene expression programming. *Neural Computing and Applications, 23*(5), 1343–1349.

Berz, G. (2000). Flood Disasters: Lessons from the pastworries for the future. *Proceedings of the Institution of Civil Engineers: Water, Maritime and Energy, 142*(1), 3–8.

Bijanvand, S., Mohammadi, M., & Parsaie, A. (2023). Estimation of water's surface elevation in compound channels with converging and diverging floodplains using soft computing techniques. *Water Supply, 23*(4), 1684–1699. https://doi.org/10.2166/ws.2023.079

Bousmar, D., Wilkin, N., Jacquemart, J. H., & Zech, Y. (2004). Overbank flow in symmetrically narrowing floodplains. *Journal of Hydraulic Engineering, 130*(4), 305–312.

Bousmar, D., & Zech, Y. (2002). Periodical turbulent structures in compound channels. In *River Flow International Conference on Fluvial Hydraulics*, Louvain-la-Neuve, Belgium, pp. 177–185.

Chlebek, J., Bousmar, D., Knight, D.W., & Sterling, M.A. (2010). Comparison of overbank flow conditions in skewed and converging/diverging channels. In: *River Flows International Conference*, pp. 503–511.

Das, B. S., Devi, K., & Khatua, K. K. (2019). Prediction of discharge in converging and diverging compound channel by gene expression programming. *Journal of Hydraulic Engineering Division of the American Society of Civil Engineers*. https://doi.org/10.1080/09715010.2018.1558116

Das, B. S., Devi, K., Khuntia, J. R., & Khatua, K. K. (2020). Discharge estimation in converging and diverging compound open channels by using adaptive neuro-fuzzy inference system. *Canadian Journal of Civil Engineering, 47*(12), 1327–1344. https://doi.org/10.1139/cjce-2018-0038

Das, B.S., Devi, K., Proust, S., & Khatua, K.K. (2018). Flow distribution in diverging compound channels using improved independent subsection method. *In River Flow 2018: 9th International Conference on Fluvial Hydraulics*, vol. 40, no. 05068.

Das, B. S., & Khatua, K. K. (2018a). Flow resistance in a compound channel with diverging and converging floodplains. *Journal of Hydraulic Engineering, 144*(8), 04018051. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001496

Das, B. S., & Khatua, K. K. (2018b). Numerical method to compute water surface profile for converging compound channel. *Arabian Journal for Science and Engineering, 43*(10), 5349–5364. https://doi.org/10.1007/s13369-018-3161-y

James, M., & Brown, R.J. (1977). Geometric parameters that influence floodplain flow. *In U.S. Army Engineer Waterways Experimental Station*, June, Vicksburg Miss. Research report H-77.

Karimi, S., Shiri, J., Kisi, O., & Shiri, A. A. (2015). Short-term and long-term streamflow prediction by using 'wavelet-gene expression' programming approach. *ISH Journal of Hydraulic Engineering, 22*(2), 148–162.

Kaushik, V., & Kumar, M. (2023a). Assessment of water surface profile in non-prismatic compound channels using machine learning techniques. *Water Supply, 23*(1), 356–378. https://doi.org/10.2166/ws.2022.430

Kaushik, V., & Kumar, M. (2023b). Sustainable gene expression programming model for shear stress prediction in nonprismatic compound channels. *Sustainable Energy Technologies and Assessments, 57*, 103229. https://doi.org/10.1016/j.seta.2023.103229

Khatua, K. K., Patra, K. C., & Mohanty, P. K. (2012). Stage-discharge prediction for straight and smooth compound channels with wide floodplains. *Journal of Hydraulic Engineering ASCE, 138*(1), 93–99.

Khuntia, J. R., Devi, K., & Khatua, K. K. (2018). Boundary shear stress distribution in straight compound channel flow using artificial neural network. *Journal of Hydrologic Engineering, 23*(5), 04018014. https://doi.org/10.1061/(asce)he.1943-5584.0001651

Knight, D. W., Tang, X., Sterling, M., Shiono, K., & McGahey, C. (2010). Solving open channel flow problems with a simple lateral distribution model. *River Flow, 1*, 41–48.

MATLAB, R. (2019). [Computer Software]. MathWorks, Natick, MA.

Myers, W. R. C., & Elsawy, E. M. (1975). Boundary shears in channel with flood plain. *J. Hydraul. Div. ASCE, 101*(7), 933–946.

Naik, B., Kaushik, V., & Kumar, M. (2022). Water surface profile in converging compound channel using gene expression programming. *Water Supply, 22*(5), 5221–5236. https://doi.org/10.2166/ws.2022.172

Naik, B., & Khatua, K. K. (2016). Water surface profile computation for compound channels with narrow flood plains. *Arabian Journal for Science and Engineering, 42*(3), 941–955. https://doi.org/10.1007/s13369-016-2236-x

Parsaie, A., & Haghiabi, A. H. (2017). Mathematical expression of discharge capacity of compound open channels using MARS technique. *Journal of Earth System Science*. https://doi.org/10.1007/s12040-017-0807-1

Parsaie, A., Najafian, S., Omid, M. H., & Yonesi, H. (2017a). Stage discharge prediction in heterogeneous compound open channel roughness. *ISH Journal of Hydraulic Engineering, 23*(1), 49–56. https://doi.org/10.1080/09715010.2016.1235471

Parsaie, A., Najafian, S., & Shamsi, Z. (2016). Predictive modeling of discharge of flow in compound open channel using radial basis neural network. *Modeling Earth Systems and Environment*. https://doi.org/10.1007/s40808-016-0207-6

Parsaie, A., Yonesi, H. A., & Najafian, S. (2015). Predictive modeling of discharge in compound open channel by support vector machine technique. *Modeling Earth Systems and Environment*. https://doi.org/10.1007/s40808-015-0002-9

Parsaie, A., Yonesi, H., & Najafian, S. (2017b). Prediction of flow discharge in compound open channels using adaptive neuro fuzzy inference system method. *Flow Measurement and Instrumentation, 54*, 288–297.

Patel, V. C. (1965). Calibration of the Preston tube and limitations on its use in pressure gradients. *Journal of Fluid Mechanics, 231*, 85–208.

Pradhan, A., & Khatua, K. K. (2017). Gene expression programming to predict Manning's n in meandering flows. *Canadian Journal of Civil Engineering, 45*(4), 304–313.

Proust, S., Rivière, N., Bousmar, D., Paquier, A., & Zech, Y. (2006). Flow in the compound channel with abrupt floodplain contraction. *Journal of Hydraulic Engineering Division of the American Society of Civil Engineers, 132*(9), 958–970.

Rezaei, B. (2006). Overbank Flow in Compound Channels with Prismatic and non-Prismatic Floodplains. Ph.D. Thesis, *University of Birmingham*, Birmingham, UK.

Rezaei, B., & Knight, D. W. (2009). Application of the Shiono and Knight Method in the compound channel with non-prismatic floodplains. *Journal of Hydraulic Research, 47*(6), 716–726.
Rezaei, B., & Knight, D. W. (2011). Overbank flow in compound channels with non-prismatic floodplains. *Journal of Hydraulic Engineering Division of the American Society of Civil Engineers, 137*, 815–824.
Sellin, R. H. J. (1964). A laboratory investigation into the interaction between flow in the channel of a river and that of its flood plain. *LaHouille Blanche, 7*, 793–801.
Yonesi, H. A., Omid, M. H., & Ayyoubzadeh, S. A. (2013). The hydraulics of flow in nonprismatic compound channels. *Journal of Civil Engineering and Urbanism, 3*(6), 342–356. https://doi.org/10.1061/(ASCE)0733-9429(2000)126:4(299)
Yonesi, H. A., Parsaie, A., Arshia, A., & Shamsi, Z. (2022). Discharge modeling in compound channels with non-prismatic floodplains using GMDH and MARS models. *Water Supply, 22*(4), 4400–4421. https://doi.org/10.2166/ws.2022.058

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

► Convenient online submission
► Rigorous peer review
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com